



# SISTEMAS EXPERTOS

# Prefacio:

La asignatura es de naturaleza teórico - práctica, tiene por objetivo que el estudiante de adquiriera conocimientos sobre las técnicas y metodologías fundamentales en el desarrollo de sistemas expertos. El estudiante desarrollara todas las habilidades para el análisis y la aplicación de diferentes métodos usados en los sistemas expertos En el campo de desarrollo de software, esta disciplina trasciende la actividad de la programación. Asimismo, pueda desarrollar las habilidades pertinentes para evaluar, en cualquier ámbito sistémico de las distintas áreas de una empresa, incluye el análisis y el diseño de sistemas informáticos.



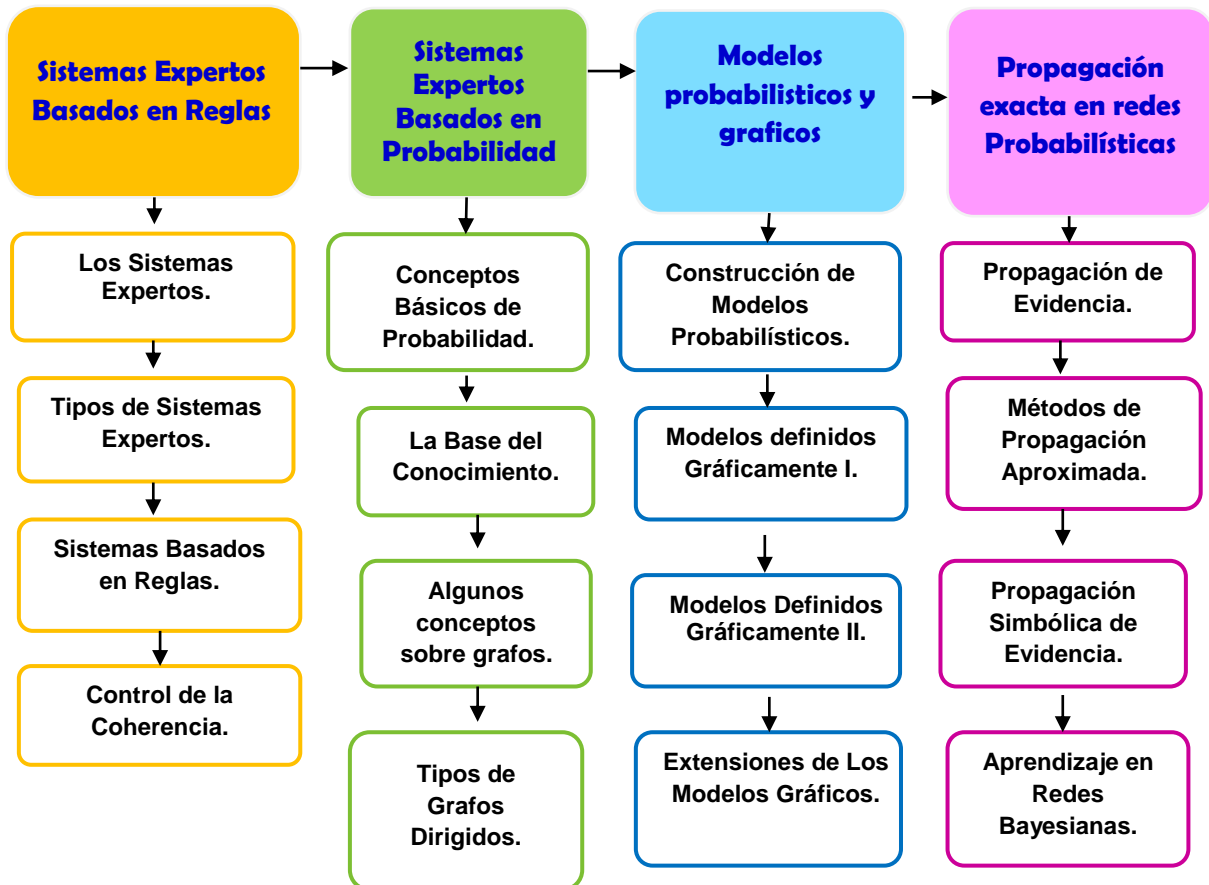
## **Comprende cuatro Unidades de Aprendizaje:**

- Unidad I: **Sistemas expertos basados en reglas.**
- Unidad II: **Sistemas expertos basados en probabilidad.**
- Unidad III: **Modelos probabilísticos y gráficos.**
- Unidad IV: **Propagación exacta en redes probabilísticas.**





## Estructura de los Contenidos



La competencia que el estudiante debe lograr al final de la asignatura es:

“Desarrollar fortalecer y perfeccionar sus habilidades en las diferentes metodologías usadas en el análisis de sistemas expertos, a través de actividades donde aplique diversas técnicas y estrategias que le permitan resolver problemas”.

# Índice del Contenido

<b>I. PREFACIO</b>	02
<b>II. DESARROLLO DE LOS CONTENIDOS</b>	03 - 156
<b>UNIDAD DE APRENDIZAJE 1: SISTEMAS EXPERTOS BASADOS EN REGLAS</b>	05-46
1. <b>Introducción</b>	06
a. Presentación y contextualización	06
b. Competencia	06
c. Capacidades	06
d. Actitudes	06
e. Ideas básicas y contenido	06
2. <b>Desarrollo de los temas</b>	07-42
a. Tema 01: Los sistemas expertos.	07
b. Tema 02: Tipos de sistemas expertos.	15
c. Tema 03: Sistemas basados en reglas.	25
d. Tema 04: Control de la coherencia.	36
3. <b>Lecturas recomendadas</b>	43
4. <b>Actividades</b>	43
5. <b>Autoevaluación</b>	44
6. <b>Resumen</b>	46
<b>UNIDAD DE APRENDIZAJE 2: SISTEMAS EXPERTOS BASADOS EN PROBABILIDAD</b>	47-86
1. <b>Introducción</b>	48
a. Presentación y contextualización	48
b. Competencia	48
c. Capacidades	48
d. Actitudes	48
e. Ideas básicas y contenido	48
2. <b>Desarrollo de los temas</b>	49-82
a. Tema 01: Conceptos básicos de probabilidad.	49
b. Tema 02: La base del conocimiento.	58
c. Tema 03: Algunos conceptos sobre grafos.	67
d. Tema 04: Tipos de grafos dirigidos.	78
3. <b>Lecturas recomendadas</b>	83
4. <b>Actividades</b>	83
5. <b>Autoevaluación</b>	84
6. <b>Resumen</b>	86
<b>UNIDAD DE APRENDIZAJE 3: MODELOS PROBABILÍSTICOS Y GRÁFICOS</b>	87-124
1. <b>Introducción</b>	88
a. Presentación y contextualización	88
b. Competencia	88
c. Capacidades	88
d. Actitudes	88
e. Ideas básicas y contenido	88
2. <b>Desarrollo de los temas</b>	89-120
a. Tema 01: Construcción de modelos probabilísticos.	89
b. Tema 02: Modelos definidos gráficamente I.	97
c. Tema 03: Modelos definidos gráficamente II.	105
d. Tema 04: Extensiones de los modelos gráficos.	112
3. <b>Lecturas recomendadas</b>	121
4. <b>Actividades</b>	121
5. <b>Autoevaluación</b>	122
6. <b>Resumen</b>	124
<b>UNIDAD DE APRENDIZAJE 4: PROPAGACIÓN EXACTA EN REDES PROBABILÍSTICAS</b>	125-153
1. <b>Introducción</b>	126
a. Presentación y contextualización	126
b. Competencia	126
c. Capacidades	126
d. Actitudes	126
e. Ideas básicas y contenido	126
2. <b>Desarrollo de los temas</b>	127-149
a. Tema 01: Propagación de evidencia.	127
b. Tema 02: Métodos de propagación aproximada.	136
c. Tema 03: Propagación simbólica de evidencia.	140
d. Tema 04: Aprendizaje en redes bayesianas.	145
3. <b>Lecturas recomendadas</b>	150
4. <b>Actividades</b>	150
5. <b>Autoevaluación</b>	151
6. <b>Resumen</b>	153
<b>III. GLOSARIO</b>	154
<b>IV. FUENTES DE INFORMACIÓN</b>	155
<b>V. SOLUCIONARIO</b>	156

# UNIDAD 1



## Sistemas Expertos Basados en Reglas



# Introducción

## **a) Presentación y contextualización**

Los temas que se tratan en la presente Unidad Temática, tienen por finalidad que el estudiante conozca el reconocimiento de la voz y el de patrones, ciertos juegos (como el ajedrez o las damas), y sistemas altamente complejos de tipo determinista o estocástico, debían ser resueltos por personas. Sin embargo, el trabajo realizado en las últimas décadas muestra que muchos de estos problemas pueden ser formulados y resueltos por maquinas basados en sistemas expertos.

## **b) Competencia**

**Aplica los fundamentos de los sistemas expertos basados en reglas para desarrollar los problemas que se presenten.**

## **c) Capacidades**

1. Reconoce los sistemas expertos en diferentes sistemas informáticos y físicos.
2. Identifica las diferentes técnicas y/ o tipos de los sistemas expertos.
3. Aplica las técnicas y herramientas de sistemas expertos basados en reglas.
4. Implementa un control de coherencias en sistemas expertos.

## **d) Actitudes**

- ✓ Presenta actitud proactiva para las soluciones de los sistemas expertos.
- ✓ Perseverancia en el desarrollo de los problemas de los sistemas expertos.

## **e) Presentación de Ideas básicas y contenido esenciales de la Unidad:**

**La Unidad de Aprendizaje 01: Sistemas expertos basados en reglas** comprende el desarrollo de los siguientes temas:

**TEMA 01: Los Sistemas Expertos.**

**TEMA 02: Tipos de Sistemas Expertos.**

**TEMA 03: Sistemas Basados en Reglas.**

**TEMA 04: Control de la Coherencia.**

# *Los Sistemas Expertos*

## TEMA 1



### Competencia:

**Reconocer los sistemas expertos en diferentes sistemas informáticos y físicos.**



# Desarrollo de los Temas



## Tema 01: Los Sistemas Expertos

El amplio campo que se conoce como inteligencia artificial (IA) trata de estos problemas, que en un principio parecían imposibles, intratables y difíciles de formular utilizando ordenadores. A. Barr y E. A. Feigenbaum, dos de los pioneros de la investigación en (IA), definen esta como sigue: “La Inteligencia Artificial



es la parte de la Ciencia que se ocupa del diseño de sistemas de computación inteligentes, es decir, sistemas que exhiben las características que asociamos a la inteligencia en el comportamiento humano que se refiere a la comprensión del lenguaje, el aprendizaje, el razonamiento, la resolución de problemas, etc.”

Hoy en día, el campo de la IA engloba varias sub áreas tales como los sistemas expertos, la demostración automática de teoremas, el juego automático, el reconocimiento de la voz y de patrones, el procesamiento del lenguaje natural, la visión artificial, la robótica, las redes neuronales, etc.

Este libro está dedicado a los sistemas expertos. Aunque los sistemas expertos constituyen una de las áreas de investigación en el campo de la IA, la mayor parte de las restantes áreas, si no todas, disponen de una componente de sistemas expertos formando parte de ellas.



### ¿QUÉ ES UN SISTEMA EXPERTO?

Los sistemas expertos son máquinas que piensan y razonan como un experto lo haría en una cierta especialidad o campo. Por ejemplo, un sistema experto en diagnóstico médico requerirá como datos los síntomas del paciente, los resultados de análisis clínicos y otros hechos relevantes, y, utilizando éstos, buscaría en una base de datos la información necesaria para poder identificar la correspondiente enfermedad.

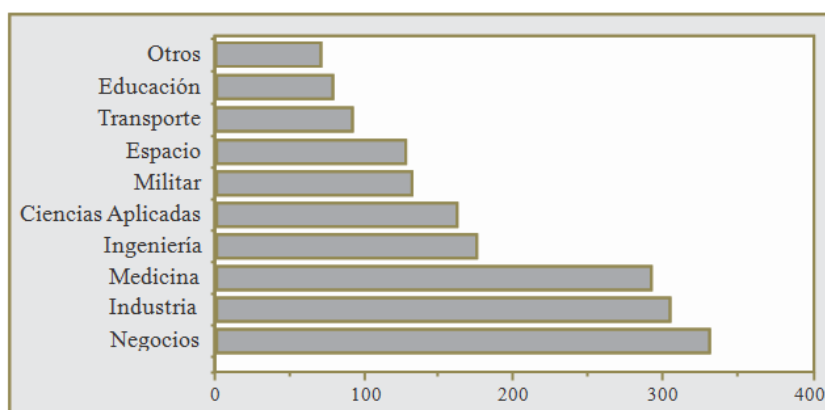


Un Sistema Experto de verdad, no solo realiza las funciones tradicionales de manejar grandes cantidades de datos, sino que también manipula esos datos de forma tal que el resultado sea inteligible y tenga significado para responder a preguntas incluso no completamente especificadas.

Aunque la anterior es todavía una definición razonable de un sistema experto, han surgido desde entonces otras definiciones, debido al rápido desarrollo de la tecnología.

**Sistema Experto:** Un sistema experto puede definirse como un sistema informático (hardware y software) que simula a los expertos humanos en un área de especialización dada. Como tal, un sistema experto deberá ser capaz de procesar y memorizar información, aprender y razonar en situaciones deterministas e inciertas, comunicar con los hombres y/u otros sistemas expertos, tomar decisiones apropiadas, y explicar por qué se han tomado tales decisiones. Se puede pensar también en un sistema experto como un consultor que puede suministrar ayuda a (o en algunos casos sustituir completamente) los expertos humanos con un grado razonable de fiabilidad.

Durante la última década se han desarrollado muy rápidamente numerosas aplicaciones de sistemas expertos a muchos campos. Durkin (1994) examina unos 2,500 sistemas expertos y los clasifica por criterios, tales como áreas de aplicación, tareas realizadas, etc. Tal como puede verse en la Figura 1.1, la economía, la industria y la medicina continúan siendo los campos dominantes entre aquellos en los que se utilizan los sistemas expertos. La sección siguiente muestra algunos ejemplos que motivan la aplicación de los sistemas expertos en algunos de estos campos.



**FIGURA 1.1. Campos de aplicación de los sistemas expertos.**

## Ejemplos Ilustrativos

Los sistemas expertos tienen muchas aplicaciones. En esta sección se dan unos pocos ejemplos ilustrativos del tipo de problemas que pueden resolverse mediante sistemas expertos. Otros ejemplos prácticos se dan a lo largo del libro.



**Ejemplo de Transacciones bancarias:** No hace mucho, para hacer una transacción bancaria, tal como depositar o sacar dinero de una cuenta, uno tenía que visitar el banco en horas de oficina. Hoy en día, esas y otras muchas transacciones pueden realizarse en cualquier momento del día o de la noche usando los cajeros automáticos que son ejemplos sencillos de sistemas expertos. De hecho, se pueden realizar estas transacciones desde casa comunicándose con el sistema experto mediante la línea telefónica.

**Ejemplo de Control de tráfico:** El control de tráfico es una de las aplicaciones más importantes de los sistemas expertos. No hace mucho tiempo, el flujo de tráfico en las calles de una ciudad se controlaba mediante guardias de tráfico que controlaban el mismo en las intersecciones. Hoy se utilizan sistemas expertos que operan automáticamente los semáforos y regulan el flujo del tráfico en las calles de una ciudad y en los ferrocarriles.

**Ejemplo de Problemas de planificación:** Los sistemas expertos pueden utilizarse también para resolver problemas complicados de planificación de forma que se optimicen ciertos objetivos como, por ejemplo, la organización y asignación de aulas para la realización de exámenes finales en una gran universidad, de forma tal que se logren los objetivos siguientes:



1. Eliminar las coincidencias de asignación simultánea de aulas: Solo se puede realizar un examen en cada aula al mismo tiempo.
2. Asientos suficientes: Un aula asignada para un examen debe tener al menos dos asientos por estudiante.
3. Minimizar los conflictos temporales: Minimizar el número de alumnos que tienen exámenes coincidentes.

4. Eliminar la sobrecarga de trabajo: Ningún alumno debe tener más de dos exámenes en un periodo de 24 horas.
5. Minimizar el número de exámenes realizados durante las tardes.

Otros ejemplos de problemas de planificación que pueden ser resueltos mediante sistemas expertos son la planificación de doctores y enfermeras en un gran hospital, la planificación en una gran factoría, y la planificación de autobuses para las horas de congestión o de días festivos.

**Ejemplo de diagnóstico médico.** Una de las aplicaciones más importantes de los sistemas expertos tiene lugar en el campo médico, donde éstos pueden ser utilizados para contestar a las siguientes preguntas:

1. ¿Cómo se puede recoger, organizar, almacenar, poner al día y recuperar la información médica (por ejemplo, registros de pacientes) de una forma eficiente y rápida? Por ejemplo, supóngase que un doctor en un centro médico está interesado en conocer información sobre una cierta enfermedad (E) y tres síntomas asociados (S1, S2, y S3). Se puede utilizar un sistema experto para buscar en la base de datos, extraer y organizar la información deseada. Esta información puede resumirse en tablas tales como la dada en la Tabla 1.1 o en gráficos como el de la Figura 1.2.
2. ¿Cómo se puede aprender de la experiencia? Es decir, como se actualiza el conocimiento de los doctores en medicina cuando el número de pacientes que éstos tratan aumenta?

3. Supuesto que un paciente presenta un conjunto de síntomas, ¿cómo se decide qué enfermedad es la que más probablemente tiene el paciente?
4. ¿Cuáles son las relaciones entre un conjunto (normalmente no observable) de enfermedades y un conjunto (observable) de síntomas? En otras palabras, ¿qué modelos pueden utilizarse para describir las relaciones entre los síntomas y las enfermedades?
5. Dado que el conjunto de síntomas conocidos no es suficiente para diagnosticar la enfermedad con cierto grado de certeza, ¿qué información adicional debe ser obtenida (por ejemplo, ¿qué síntomas adicionales deben ser identificados? o ¿qué pruebas médicas deben realizarse?).

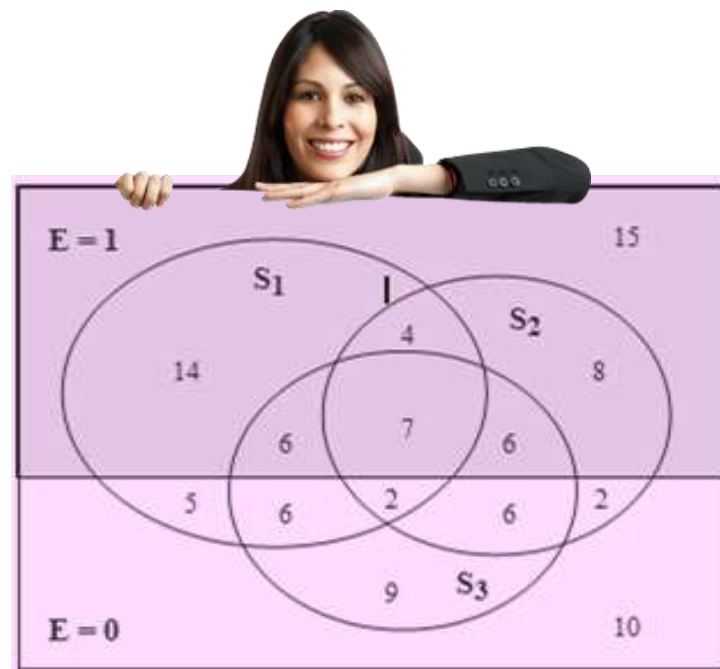
6. ¿Cuál es el valor de cada una de estas piezas de información? En otras palabras, ¿cuál es la contribución de cada uno de los síntomas adicionales o pruebas a la toma de decisión?

**Ejemplo de Agentes secretos.** Alberto, Luisa, Carmen, y Tomas son agentes secretos, cada uno está en uno de los cuatro países: Egipto, Francia, Japón y España. No se sabe dónde está cada uno de ellos. Por tanto, se ha pedido información y se han recibido los cuatro telegramas siguientes:



- ❖ Desde Francia: Luisa está en España.
- ❖ Desde España: Alberto está en Francia.
- ❖ Desde Egipto: Carmen está en Egipto.
- ❖ Desde Japón: Carmen está en Francia.

No se sabe quién es el que ha mandado cada uno de los mensajes, pero se sabe que Tomas miente (¿un agente doble?) y que los demás agentes dicen la verdad.



**FIGURA 1.2.** Una representación gráfica de la distribución de frecuencias de una enfermedad (D) y tres síntomas binarios (S1, S2, y S3) en una base de datos médica.



$E$	$S_1$	$S_2$	$S_3$	Frecuencia
1	1	1	1	7
1	1	1	0	4
1	1	0	1	6
1	1	0	0	14
1	0	1	1	6
1	0	1	0	8
1	0	0	1	0
1	0	0	0	15
0	1	1	1	2
0	1	1	0	0
0	1	0	1	6
0	1	0	0	5
0	0	1	1	6
0	0	1	0	2
0	0	0	1	9
0	0	0	0	10



**TABLA 1.1.** Una representación tabular de la distribución de frecuencias de una enfermedad ( $D$ ) y tres síntomas binarios ( $S_1$ ,  $S_2$ , y  $S_3$ ) en una base de datos médica (1 representa la presencia y 0 representa la ausencia de la enfermedad o el síntoma indicado).



### ¿POR QUÉ LOS SISTEMAS EXPERTOS?

El desarrollo o la adquisición de un sistema experto es generalmente caro, pero el mantenimiento y el coste marginal de su uso repetido es relativamente bajo. Por otra parte, la ganancia en términos monetarios, tiempo, y precisión resultantes del uso de los sistemas expertos son muy altas, y la amortización es muy rápida. Sin embargo, antes de desarrollar o adquirir un sistema experto debe realizarse un análisis de factibilidad y de coste-beneficio.

**Hay varias razones para utilizar sistemas expertos. Las más importantes son:**

1. Con la ayuda de un sistema experto, personal con poca experiencia puede resolver problemas que requieren un conocimiento de experto. Esto es también importante en casos en los que hay pocos expertos humanos. Además, el número de personas con acceso al conocimiento aumenta con el uso de sistemas expertos.

2. El conocimiento de varios expertos humanos puede combinarse, lo que da lugar a sistemas expertos más fiables, ya que se obtiene un sistema experto que combina la sabiduría colectiva de varios expertos humanos en lugar de la de uno solo.

3. Los sistemas expertos pueden responder a preguntas y resolver problemas mucho más rápidamente que un experto humano. Por ello, los sistemas son muy valiosos en casos en los que el tiempo de respuesta es crítico.



4. En algunos casos, la complejidad del problema impide al experto humano resolverlo. En otros casos la solución de los expertos humanos no es fiable. Debido a la capacidad de los ordenadores de procesar un elevadísimo número de operaciones complejas de forma rápida y aproximada, los sistemas expertos suministran respuestas rápidas y fiables en situaciones en las que los expertos humanos no pueden.

5. Los sistemas expertos pueden ser utilizados para realizar operaciones monótonas, aburridas e incómodas para los humanos. En verdad, los sistemas expertos pueden ser la única solución viable en una situación en la que la tarea a realizar desborda al ser humano (por ejemplo, un avión o una capsula espacial dirigida por un sistema experto).

6. Se pueden obtener enormes ahorros mediante el uso de sistemas expertos. El uso de los sistemas expertos se recomienda especialmente en las situaciones siguientes:

7. Cuando el conocimiento es difícil de adquirir o se basa en reglas que solo pueden ser aprendidas de la experiencia.

8. Cuando la mejora continua del conocimiento es esencial y/o cuando el problema está sujeto a reglas o códigos cambiantes.

9. Cuando los expertos humanos son caros o difíciles de encontrar.

10. Cuando el conocimiento de los usuarios sobre el tema es limitado.

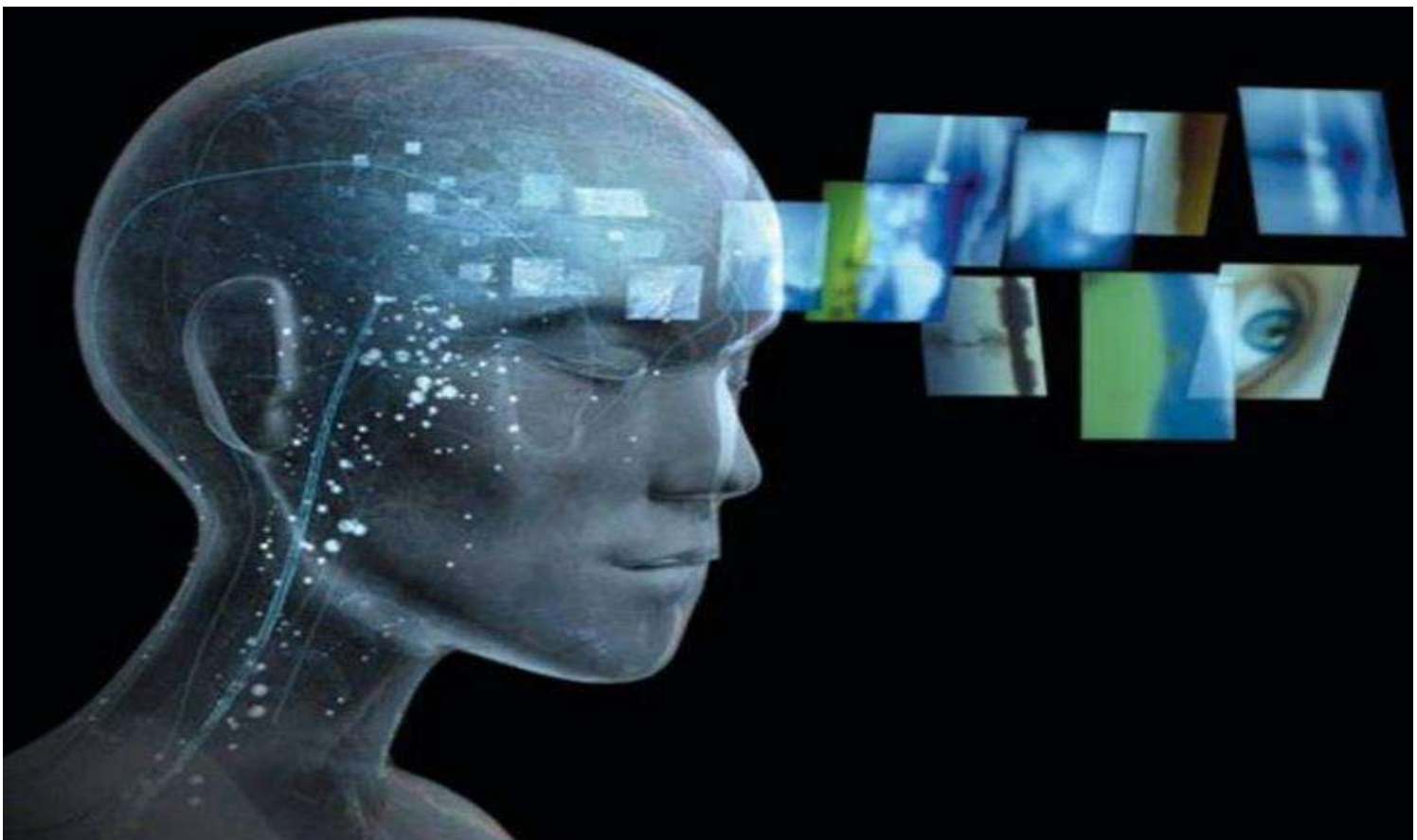
# *Tipos de Sistemas Expertos*

## TEMA 2



### Competencia:

**Identificar las diferentes técnicas y/ o tipos de los sistemas expertos.**





## Tema 02: Tipos de Sistemas Expertos

Los problemas con los que pueden tratar los sistemas expertos pueden clasificarse en dos tipos: problemas esencialmente deterministas y problemas esencialmente estocásticos. Consecuentemente, los sistemas expertos pueden clasificarse en dos tipos principales según la naturaleza de problemas para los que están diseñados:



deterministas y estocásticos. Los problemas de tipo determinista pueden ser formulados usando un conjunto de reglas que relacionen varios objetos bien definidos. Los sistemas expertos que tratan problemas deterministas son conocidos como sistemas basados en reglas, porque sacan sus conclusiones basándose en un conjunto de reglas utilizando un mecanismo de razonamiento lógico.

En situaciones inciertas, es necesario introducir algunos medios para tratar la incertidumbre. Por ejemplo, algunos sistemas expertos usan la misma estructura de los sistemas basados en reglas, pero introducen una medida asociada a la incertidumbre de las reglas y a la de sus premisas.

En este caso se pueden utilizar algunas fórmulas de propagación para calcular la incertidumbre asociada a las conclusiones. Durante las últimas décadas han sido propuestas algunas medidas de incertidumbre.



Algunos ejemplos de estas medidas son los factores de certeza, usados en las conchas para generar sistemas expertos tales como el sistema experto MYCIN; la lógica difusa y la teoría de la evidencia de Dempster y Shafer.

Otra medida intuitiva de incertidumbre es la probabilidad, en la que la distribución conjunta de un conjunto de variables se usa para describir las relaciones de dependencia entre ellas, y se sacan conclusiones usando formulas muy conocidas de la teoría de la

probabilidad.

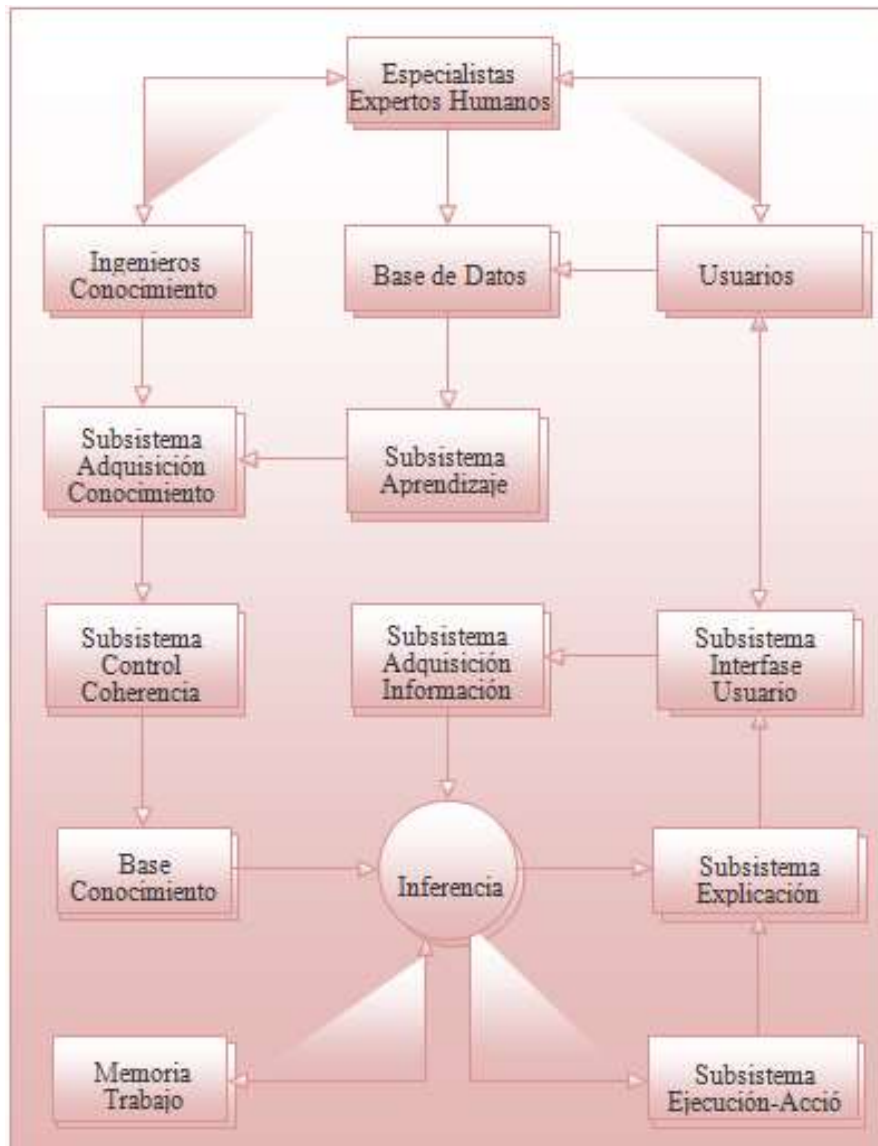


Este es el caso del sistema experto PROSPECTOR, que utiliza el teorema de Bayes para la exploración de mineral. Los sistemas expertos que utilizan la probabilidad como medida de incertidumbre se conocen como sistemas expertos probabilísticos y la estrategia de razonamiento que usan se conoce como razonamiento probabilístico, o inferencia probabilística. Este libro está dedicado a los sistemas expertos de tipo probabilístico.

En los comienzos de los sistemas expertos de tipo probabilístico surgieron varios obstáculos, debido a las dificultades encontradas para definir la distribución de probabilidad conjunta de las variables. Ello ha ralentizado su desarrollo. Con la introducción de los modelos de redes probabilísticas, estos obstáculos se han superado y los sistemas expertos probabilísticos han vuelto de forma espectacular durante las dos últimas décadas.

Estos modelos, que incluyen las redes de Markov y las Bayesianas, se basan en una representación gráfica de las relaciones entre las variables. Esta representación conduce no solo a formas más eficientes de definir la distribución conjunta de probabilidad sino también a una propagación de incertidumbre muy eficiente, que permite sacar conclusiones. Componentes de un Sistema Experto Las definiciones de sistemas expertos dadas se entienden mejor cuando se examinan las principales componentes de los sistemas expertos. Estas componentes se muestran esquemáticamente en la Figura 1.3 y se explican seguidamente.





**Figura 1.3 Componentes típicos de un sistema experto.**

## Componente Humano

Un sistema experto es generalmente el resultado de la colaboración de uno o varios expertos humanos especialistas en el tema de estudio y los ingenieros del conocimiento, con los usuarios en mente. Los expertos humanos suministran el conocimiento básico en el tema de interés, y los ingenieros del conocimiento trasladan este conocimiento a un lenguaje, que el sistema experto pueda entender. La colaboración de los expertos humanos, los ingenieros del conocimiento y los usuarios es, quizás, el elemento más importante en el desarrollo de un sistema experto. Esta etapa requiere una enorme dedicación y un gran esfuerzo debido a los diferentes lenguajes que hablan las distintas partes y a las diferentes experiencias que tienen.

## LA BASE DE CONOCIMIENTO

Los especialistas son responsables de suministrar a los ingenieros del conocimiento una base de conocimiento ordenada y estructurada, y un conjunto de relaciones bien definidas y explicadas. Esta forma estructurada de pensar requiere que los expertos humanos repiensen, reorganicen, y reestructuren la base de conocimiento y, como resultado, el especialista se convierte en un mejor conocedor de su propio campo de especialidad. Hay que diferenciar entre datos y conocimiento. El conocimiento se refiere a afirmaciones de validez general tales como reglas, distribuciones de probabilidad, etc. Los datos se refieren a la información relacionada con una aplicación particular.

Por ejemplo, en diagnóstico médico, los síntomas, las enfermedades y las relaciones entre ellos, forman parte del conocimiento, mientras los síntomas particulares de un paciente dado forman parte de los datos. Mientras el conocimiento es permanente, los datos son efímeros, es decir, no forman parte de la componente permanente de un sistema y son destruidos después de usarlos. El conocimiento se almacena en la base de Conocimiento y los datos se almacenan en la memoria de trabajo. Todos los Procedimientos de los diferentes sistemas y subsistemas que son de carácter transitorio se almacenan también en la memoria de trabajo.

## SUBSISTEMA DE ADQUISICIÓN DE CONOCIMIENTO

El subsistema de adquisición de conocimiento controla el flujo del nuevo conocimiento que fluye del experto humano a la base de datos. El sistema determina que nuevo conocimiento se necesita, o si el conocimiento recibido es en realidad nuevo, es decir, si debe incluirse en la base de datos y, en caso necesario, incorpora estos conocimientos a la misma.

### Control de la Coherencia

El subsistema de control de la coherencia ha aparecido en los sistemas expertos muy recientemente. Sin embargo, es una componente esencial de un sistema experto. Este subsistema controla la consistencia de la base de datos y evita que unidades de conocimiento inconsistentes entren en la misma. En situaciones complejas incluso un experto humano puede formular afirmaciones inconsistentes.

Por ello, sin un subsistema de control de la coherencia, unidades de conocimiento contradictorio pueden formar parte de la base de conocimiento, dando lugar a un comportamiento insatisfactorio del sistema. Es también bastante común, especialmente en sistemas con mecanismos de propagación de incertidumbre, que se llegue a conclusiones absurdas o en conflicto como, por ejemplo, situaciones en las que el sistema genera probabilidades mayores que la unidad o negativas. Por ello, el subsistema de control de la coherencia comprueba e informa a los expertos de las inconsistencias. Por otra parte, cuando se solicita información de los expertos humanos, éste subsistema informa sobre las restricciones que ésta debe cumplir para ser coherente con la existente en la base de conocimiento. De esta forma, ayuda a los expertos humanos a dar información fiable.



## MOTOR DE INFERENCIA

El motor de inferencia es el corazón de todo sistema experto. El cometido principal de esta componente es el de sacar conclusiones aplicando el conocimiento a los datos. Por ejemplo, en diagnóstico médico, los síntomas de un paciente (datos) son analizados a la luz de los síntomas y las enfermedades y de sus relaciones (conocimiento). Las conclusiones del motor de inferencia pueden estar basadas en conocimiento determinista o conocimiento probabilístico. Como puede esperarse, el tratamiento de situaciones de incertidumbre (probabilísticas) puede ser considerablemente más difícil que el tratamiento de situaciones ciertas (deterministas).



En muchos casos, algunos hechos (datos) no se conocen con absoluta certeza. Por ejemplo, piénsese en un paciente que no está seguro de sus síntomas. Puede darse el caso de tener que trabajar con conocimiento de tipo no determinista, es decir, de casos en los que se dispone solo de información aleatoria o difusa. El motor de inferencia es también responsable de la propagación de este conocimiento incierto. De hecho, en los sistemas expertos basados en probabilidad, la propagación de incertidumbre es la tarea principal del motor de inferencia, que permite sacar conclusiones bajo incertidumbre. Esta tarea es tan compleja que da lugar a que ésta sea probablemente la componente más débil de casi todos los sistemas expertos existentes. Por esta razón, la mayor parte de este libro se dedica al análisis y resolución del problema de la propagación de incertidumbre.



### **EL SUBSISTEMA DE ADQUISICIÓN DE CONOCIMIENTO**

Si el conocimiento inicial es muy limitado y no se pueden sacar conclusiones, el motor de inferencia utiliza el subsistema de adquisición de conocimiento para obtener el conocimiento necesario y continuar con el proceso de inferencia hasta que se hayan sacado conclusiones. En algunos casos, el usuario puede suministrar la información requerida para éste y otros objetivos. De ello resulta la necesidad de una interface de usuario y de una comprobación de la consistencia de la información suministrada por el usuario antes de introducirla en la memoria de trabajo.

### **INTERFACE DE USUARIO**

La interface de usuario es el enlace entre el sistema experto y el usuario. Por ello, para que un sistema experto sea una herramienta efectiva, debe incorporar mecanismos eficientes para mostrar y obtener información de forma fácil y agradable. Un ejemplo de la información que tiene que ser mostrada tras el trabajo del motor de inferencia, es el de las conclusiones, las razones que expliquen tales conclusiones y una explicación de las acciones iniciadas por el sistema experto. Por otra parte, cuando el motor de inferencia no puede concluir debido, por ejemplo, a la ausencia de información, la interface de usuario es un vehículo para obtener la información necesaria del usuario. Consecuentemente, una implementación inadecuada de la interface de usuario que no facilite este proceso minaría notablemente la calidad de un sistema experto.

Otra razón de la importancia de la interface de usuario es que los usuarios evalúan comúnmente los sistemas expertos y otros sistemas por la calidad de dicha interface más que por la del sistema experto mismo, aunque no se debería juzgar la calidad de un libro por su portada.

### **El Subsistema de Ejecución de Ordenes**

El subsistema de ejecución de órdenes es la componente que permite al sistema experto iniciar acciones. Estas acciones se basan en las conclusiones sacadas por el motor de inferencia. Como ejemplos, un sistema experto diseñado para analizar el tráfico ferroviario puede decidir retrasar o parar ciertos trenes para optimizar el tráfico global, o un sistema para controlar una central nuclear puede abrir o cerrar ciertas válvulas, mover barras, etc., para evitar un accidente. La explicación de las razones por las que se inician estas acciones pueden darse al usuario mediante el subsistema de explicación.



## EL SUBSISTEMA DE EXPLICACIÓN

El usuario puede pedir una explicación de las conclusiones sacadas o de las acciones iniciadas por el sistema experto. Por ello, es necesario un subsistema que explique el proceso seguido por el motor de inferencia o por el subsistema de ejecución. Por ejemplo, si un cajero automático decide

rechazar la palabra clave (una acción), la maquina puede mostrar un mensaje (una explicación) como la siguiente:

¡Lo siento!, su palabra clave es todavía incorrecta tras tres intentos.

Retenemos su tarjeta de crédito, para garantizar su seguridad.

Por favor, póngase en contacto con su banco en horas de oficina

En muchos dominios de aplicaciones, es necesaria la explicación de las conclusiones debido a los riesgos asociados con las acciones a ejecutar. Por ejemplo, en el campo del diagnóstico médico, los doctores son responsable ´últimos de los diagnósticos, independientemente de las herramientas técnicas utilizadas para sacar conclusiones. En estas situaciones, sin un subsistema de explicación, los doctores pueden no ser capaces de explicar a sus pacientes las razones de su diagnóstico.

## El Subsistema de Aprendizaje

Una de las principales características de un sistema experto es su capacidad para aprender. Diferenciaremos entre aprendizaje estructural y aprendizaje paramétrico. Por aprendizaje estructural nos referimos a algunos aspectos relacionados con la estructura del conocimiento (reglas, distribuciones de probabilidad, etc.).



Por ello, el descubrimiento de nuevos síntomas relevantes para una enfermedad o la inclusión de una nueva regla en la base de conocimiento son ejemplos de aprendizaje estructural. Por aprendizaje paramétrico nos referimos a estimar los parámetros necesarios para construir la base de conocimiento. Por ello, la estimación de frecuencias o probabilidades asociadas a síntomas o enfermedades es un ejemplo de aprendizaje paramétrico.



Otra característica de los sistemas expertos es su habilidad para obtener experiencia a partir de los datos disponibles. Estos datos pueden ser obtenidos por expertos y no expertos y pueden utilizarse por el subsistema de adquisición de conocimiento y por el subsistema de aprendizaje. De las componentes antes mencionadas puede verse que los sistemas expertos pueden realizar varias tareas.

## Estas tareas incluyen, pero no se limitan a, las siguientes:

- ✚ Adquisición de conocimiento y la verificación de su coherencia; por lo que el sistema experto puede ayudar a los expertos humanos a dar conocimiento coherente.
- ✚ Almacenar (memorizar) conocimiento.
- ✚ Preguntar cuando se requiere nuevo conocimiento.
- ✚ Realizar inferencia y razonamiento en situaciones deterministas y de incertidumbre.
- ✚ Explicar conclusiones o acciones tomadas.
- ✚ Comunicar con los expertos y no expertos humanos y con otros sistemas expertos.



## Desarrollo de un Sistema Experto

Weiss y Kulikowski (1984) sugieren las etapas siguientes para el diseño e implementación de un sistema experto, Figura 1.4.

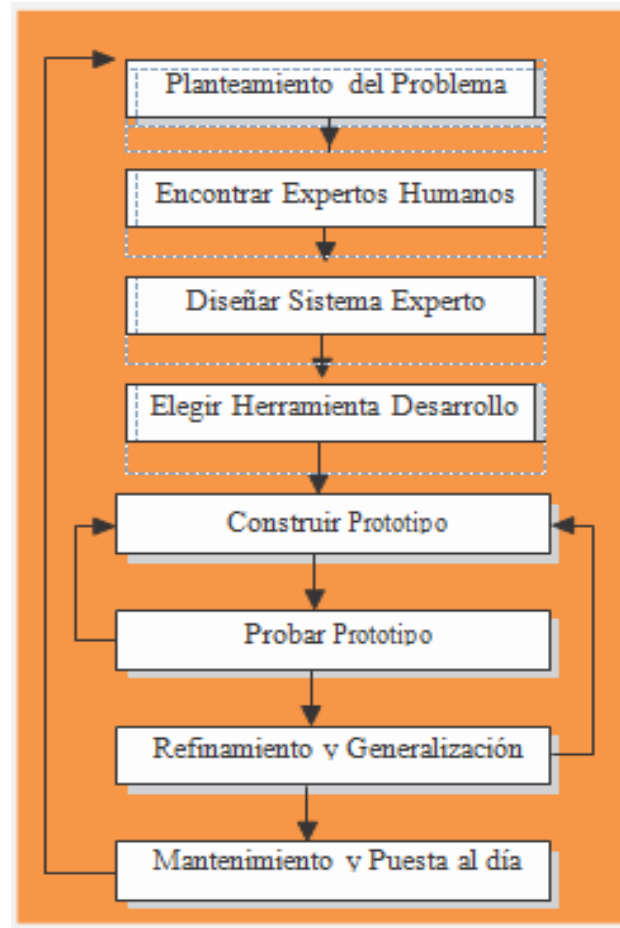


Figura 1.4 Etapas en el desarrollo de un sistema experto.

### Planteamiento del problema.

La primera etapa en cualquier proyecto es normalmente la definición del problema a resolver. Puesto que el objetivo principal de un sistema experto es responder a preguntas y resolver problemas, esta etapa es quizás la más importante en el desarrollo de un sistema experto. Si el sistema está mal definido, se espera que el sistema suministre respuestas erróneas.

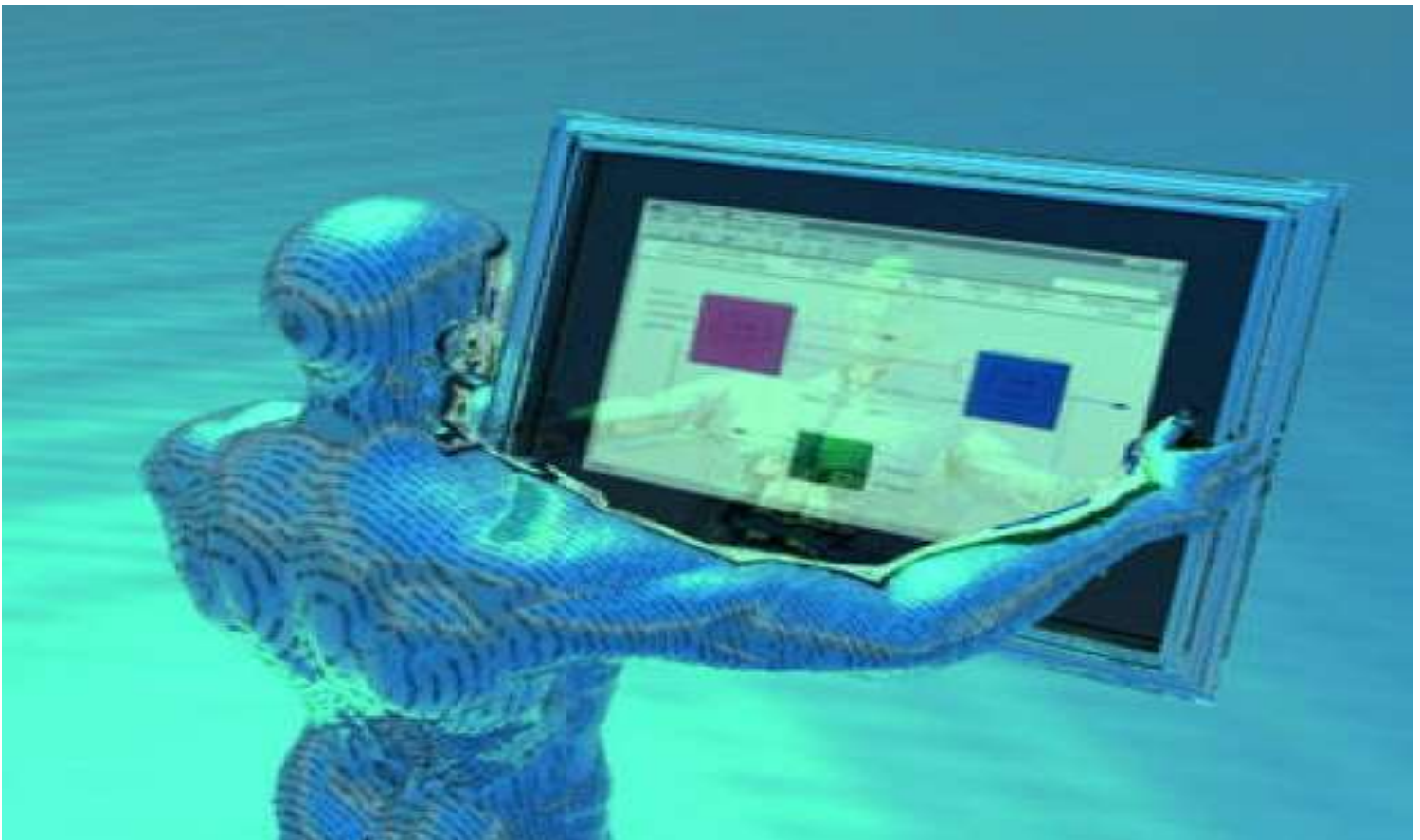
# *Sistemas Basados en Reglas*

## TEMA 3



### Competencia:

**Aplicar las técnicas y herramientas de sistemas expertos basados en reglas.**







## Tema 03: Sistemas Basados en Reglas



En nuestra vida diaria encontramos muchas situaciones complejas gobernadas por reglas deterministas: sistemas de control de tráfico, sistemas de seguridad, transacciones bancarias, etc. Los sistemas basados en reglas son una herramienta eficiente para tratar estos problemas. Las reglas deterministas constituyen la más sencilla de las metodologías utilizadas en sistemas expertos. La base de conocimiento contiene el conjunto de reglas que definen el problema, y el motor de inferencia saca las conclusiones aplicando la lógica clásica a estas reglas.

El libro de Pedersen (1989) muestra un enfoque práctico e incluye varios algoritmos. Describe la base de conocimiento de los sistemas expertos basados en reglas y da una definición y ejemplos de reglas, que constituyen el corazón de la base de conocimiento. Seguidamente, se discute como opera el motor de inferencia.

Objeto	Conjunto de valores posibles
Nota Calificación	{0, 1, ..., 10}
Puesto Admitir	{sobresaliente, notable, aprobado, suspenso}
Notificar	{0, 1, ..., 100}
	{sí, pendiente, no}
	{sí, no}

Tabla 2.1 Un ejemplo de objetos con sus posibles valores.

### LA BASE DE CONOCIMIENTO

En los sistemas basados en reglas intervienen dos elementos importantes: la base de conocimiento y los datos. Los datos están formados por la evidencia o los hechos conocidos en una situación particular. Este elemento es dinámico, es decir, puede cambiar de una aplicación a otra. Por esta razón, no es de naturaleza permanente y se almacena en la memoria de trabajo. En situaciones deterministas, las relaciones entre un conjunto de objetos pueden ser representadas mediante un conjunto de reglas.





El conocimiento se almacena en la base de conocimiento y consiste en un conjunto de objetos y un conjunto de reglas que gobiernan las relaciones entre esos objetos. La información almacenada en la base de conocimiento es de naturaleza permanente y estática, es decir, no cambia de una aplicación a otra, a menos que se incorporen al sistema experto elementos de

aprendizaje. Para dar una idea intuitiva de lo que es una regla, supóngase que se tiene un conjunto de *objetos* y, por simplicidad, que cada objeto puede tener uno y solo uno de un conjunto de posibles valores. Ejemplos de objetos con sus posibles valores se dan en la Tabla 2.1. Seguidamente se dan unos pocos ejemplos de reglas:

**Regla 1:** Si  $\text{nota} > 9$ , entonces calificación = sobresaliente.

**Regla 2:** Si  $\text{puesto} < 20$  o  $\text{nota} > 7$ , entonces Admitir = sí y Notificar = sí.

Cada una de las reglas anteriores relaciona dos o más objetos y está formada por las partes siguientes:

- La premisa de la regla, que es la expresión lógica entre las palabras clave *si* y *entonces*. La premisa puede contener una o más afirmaciones objeto-valor conectadas con operadores lógicos **y**, **o**, o **no**. Por ejemplo, la premisa de la Regla 1 consta de una única afirmación objeto-valor, mientras que las premisas de la Regla 2 constan de dos afirmaciones objeto-valor conectadas por un operador lógico.
- La *conclusión* de la regla, que es la expresión lógica tras la palabra clave *entonces*.

**REGLA.** Una regla es una afirmación lógica que relaciona dos o más objetos e incluye dos partes, la premisa y la conclusión. Cada una de estas partes consiste en una expresión lógica con una o más afirmaciones objeto-valor conectadas mediante los operadores lógicos **y**, **o**, o **no**. Una regla se escribe normalmente como “Si premisa, entonces conclusión”. En general, ambas, la premisa y la conclusión de una regla, pueden contener afirmaciones múltiples objeto-valor.



Una expresión lógica que contiene solo una afirmación objeto-valor se denomina expresión lógica simple; en caso contrario, la expresión se dice expresión lógica compuesta. Por ejemplo, las expresiones lógicas en ambas, premisa y conclusión de la Regla 1, son simples, mientras que las expresiones lógicas de las premisas y la conclusión de la Regla 2 es compuesta. Correspondientemente, una regla que contiene solamente expresiones lógicas simples se denomina una regla simple; en otro caso, se llama regla compuesta. Por ejemplo, la Regla 1 es simple, mientras que la Reglas 2 es compuesta.

**CAJERO AUTOMÁTICO.** Como ejemplo de problema determinista que puede ser formulado usando un conjunto de reglas, considérese una situación en la que un usuario (por ejemplo, un cliente) desea sacar dinero de su cuenta corriente mediante un cajero automático (CA). En cuanto el usuario introduce la tarjeta en el CA, la maquina la lee y la verifica. Si la tarjeta no es verificada con éxito (por ejemplo, porque no es legible), el CA devuelve la tarjeta al usuario con el mensaje de error correspondiente. En otro caso, el CA pide al usuario su número de identificación personal (NIP). Si el número fuese incorrecto, se dan tres oportunidades de corregirlo. Si el NIP es correcto, el CA pregunta al usuario cuánto dinero desea sacar. Para que el pago se autorice, la cantidad solicitada no debe exceder de una cierta cantidad límite diaria, además de haber suficiente dinero en su cuenta.

En este caso se tienen siete objetos, y cada objeto puede tomar uno y solo un valor de entre sus posibles valores. La Tabla muestra estos objetos y sus posibles valores. La Figura muestra siete reglas que gobiernan la estrategia que el CA debe seguir cuando un usuario intenta sacar dinero de su cuenta. En la Regla 1, por ejemplo, la premisa consiste en seis afirmaciones objeto - valor conectado mediante el operador lógico y, lo que indica que la premisa es cierta si las seis afirmaciones lo son. Por ello, la Regla 1 relaciona el objeto *Pago* (en la conclusión) con los demás objetos. Según la Regla 1, la acción que debe iniciar el CA es dar el dinero al usuario si la tarjeta se ha verificado correctamente, la fecha no ha expirado, el NIP es correcto, el número de intentos para dar el NIP correcto no se ha excedido y la cantidad solicitada no excede ni la cantidad disponible ni el límite máximo diario. Las expresiones lógicas en cada una de las restantes reglas de la Figura 1.5 constan de una sola afirmación. Nótese que la Regla 1 indica cuando debe permitirse el pago, y las restantes cuando debe rechazarse.

Objeto	Conjunto de posibles valores
Tarjeta	{verificada, no verificada}
Fecha	{expirada, no expirada}
NIP	{correcto, incorrecto}
Intentos	{excedidos, no excedidos}
Balance	{suficiente, insuficiente}
Límite	{excedido, no excedido}
Pago	{autorizado, no autorizado}

Tabla 2.2 Objetos y posibles valores para el ejemplo del cajero automático.

**Gente famosa:** Supóngase que se dispone de una base de datos consistente en  $N$  individuos. Para cada individuo, la base de datos contiene cuatro atributos: nombre, sexo, nacionalidad y profesión. Supóngase que la base de datos muestra solo si una persona es americana, política y/o si es mujer. Cada uno de estos atributos es binario (toma solo dos valores posibles). En este caso, la base de datos puede contener, como mucho,  $2^3 = 8$  conjuntos disjuntos. Estos conjuntos se muestran en la Figura 1.6. La figura muestra también el nombre de una persona en cada subconjunto. La Tabla 2.3 da un ejemplo de una base de datos que contiene  $N = 8$  personas famosas. En este caso se tienen cuatro objetos: *Nombre*, *americano*,



**Político, y Mujer.** El primer objeto puede tomar uno de  $N$  posibles valores (los nombres de cada persona) y cada uno de los tres últimos objetos pueden tomar el valor *sí* o el valor *no*. A partir de la Tabla 2.3 se pueden construir reglas para identificar a cada persona, resultando un total de ocho reglas. Por ejemplo, la regla siguiente corresponde al presidente Clinton:

**Regla 1:** Si *Nombre = Clinton*, entonces *Americano = sí* y *Político = sí* y *Mujer = no*. Las restantes siete reglas pueden construirse de forma análoga.

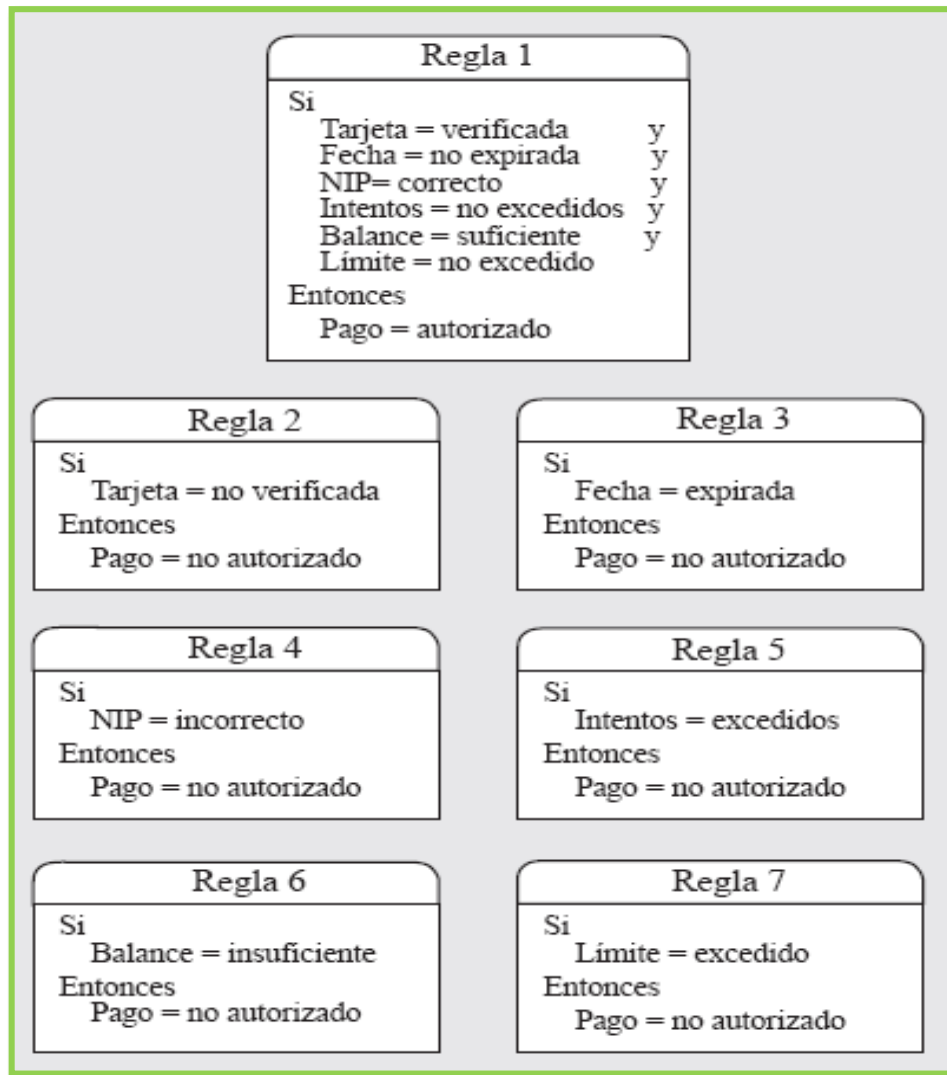


FIGURA 1.5. Ejemplos de reglas para sacar dinero de un *cajero automático*.

Nombre	American	Político	Mujer
Barbara Jordan	sí	sí	sí
Bill Clinton	sí	sí	no
Barbara Walters	sí	no	sí
Mohammed Ali	sí	no	no
Margaret	no	sí	sí
Thatcher Anwar	no	sí	no
El-Sadat Marie	no	no	sí
Curie	no	no	no
Pablo Picasso			

TABLA 2.3. Una base de datos mostrando cuatro objetos y sus valores correspondientes para el ejemplo de las personas famosas.



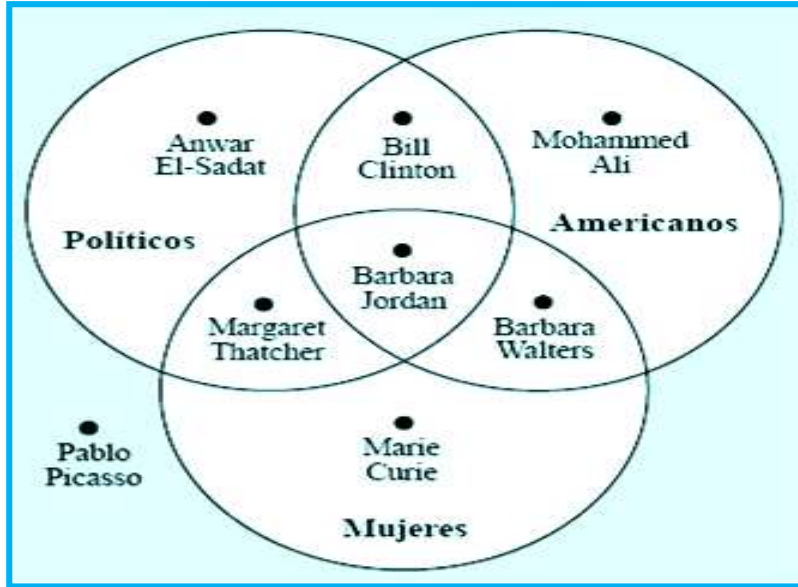


FIGURA 1.6. Un ejemplo de una base de datos con tres atributos binarios que dividen la población en ocho conjuntos disjuntos.

### Algunos sistemas imponen ciertas restricciones a las reglas. Por ejemplo:

- No permitir en la premisa el operador lógico o, y
- Limitar las conclusiones a expresiones lógicas simples.

Hay buenas razones para imponer estas restricciones. En primer lugar, las reglas que satisfacen estas restricciones son fáciles de tratar a la hora de escribir un programa de ordenador. En segundo lugar, las dos restricciones anteriores no dan lugar a una pérdida de generalidad, puesto que reglas mucho más generales pueden ser reemplazadas por conjuntos de reglas de esta forma. A esto se le llama *sustitución de reglas*. Por tanto, el conjunto de reglas especificado inicialmente por el experto humano puede requerir una sustitución posterior por un conjunto de reglas equivalente para satisfacer estas restricciones.



La Tabla 2.4 da ejemplos de sustitución de reglas. Nótese que cada regla de la primera columna puede ser sustituida por el correspondiente conjunto de reglas de la segunda columna y que todas las reglas de ésta satisfacen las condiciones anteriores. Por ejemplo, la primera regla compuesta de la Tabla 2.4: • Regla 1: Si A o B, entonces C, Puede ser reemplazada por las dos reglas simples:

Regla	Reglas Equivalentes
Si $A$ o $B$ , entonces $C$	Si $A$ , entonces $C$ Si $B$ , entonces $C$
Si $A$ o $B$ , entonces $C$	Si $\bar{A}$ y $\bar{B}$ , entonces $C$
Si $A$ y $B$ , entonces $C$	Si $\bar{A}$ , entonces $C$ Si $\bar{B}$ , entonces $C$
Si $(A$ o $B)$ y $C$ , entonces $D$	Si $A$ y $C$ , entonces $D$ Si $B$ y $C$ , entonces $D$
Si $(A$ o $B)$ y $C$ , entonces $D$	Si $\bar{A}$ y $\bar{B}$ y $C$ , entonces $D$
Si $A$ y $B$ y $C$ , entonces $D$	Si $\bar{A}$ y $C$ , entonces $D$ Si $\bar{B}$ y $C$ , entonces $D$
Si $A$ , entonces $B$ y $C$	Si $A$ , entonces $B$ Si $A$ , entonces $C$
Si $A$ , entonces $B$ o $C$	Si $A$ y $\bar{B}$ , entonces $C$ Si $A$ y $C$ , entonces $B$
Si $A$ , entonces $B$ y $C$	Si $A$ y $B$ , entonces $\bar{C}$ Si $A$ y $C$ , entonces $\bar{B}$
Si $A$ , entonces $B$ o $C$	Si $A$ , entonces $\bar{B}$ Si $A$ , entonces $C$

**TABLA 2.4.** Ejemplos de sustitución de reglas: Las reglas en la primera columna son equivalentes a las reglas de la segunda columna. Nótese que en los seis primeros ejemplos las sustituciones se aplican a la premisa y en los cuatro últimos, a la conclusión.

- Regla 1a: Si  $A$ , entonces  $C$ .
- Regla 1b: Si  $B$ , entonces  $C$ .

**Como ejemplo adicional, la Tabla 2.5 muestra que:**

- Regla 2: Si  $A$  o  $B$ , entonces  $C$ , puede ser reemplazada por la regla
- Regla 2: Si  $\bar{A}$  y  $\bar{B}$ , entonces  $C$ , donde  $\bar{A}$  significa *no A*. La Tabla 2.5 se llama *tabla de verdad*.

	$B$	$\bar{A}$	$\bar{B}$	$A$ o $B$	$\bar{A}$ y $\bar{B}$
C	C	F	F	F	F
C	F	F	C	F	F
F	C	C	F	F	F
F	F	C	C	C	C

**TABLA 2.5.** Una tabla de verdad mostrando que las expresiones lógicas  $A$  o  $B$  y  $\bar{A}$  y  $\bar{B}$  son equivalentes. Los símbolos  $C$  y  $F$  se utilizan para cierto y falso, respectivamente.

## EL MOTOR DE INFERENCIA

Tal como se ha mencionado en la sección anterior, hay dos tipos de elementos: los datos (hechos o evidencia) y el conocimiento (el conjunto de reglas almacenado en la base de conocimiento). El motor de inferencia usa ambos para obtener nuevas conclusiones o hechos. Por ejemplo, si la premisa de una regla es cierta, entonces la conclusión de la regla debe ser también cierta. Los datos iniciales se incrementan incorporando las nuevas conclusiones. Por ello, tanto los hechos iniciales o datos de partida como las conclusiones derivadas de ellos forman parte de los hechos o datos de que se dispone en un instante dado.



### Las conclusiones pueden clasificarse en dos tipos: *simples* y *compuestas*.

Las conclusiones simples son las que resultan de una regla simple. Las conclusiones compuestas son las que resultan de más de una regla. Para obtener conclusiones, los expertos utilizan diferentes tipos de reglas y estrategias de inferencia. En el resto de esta sección se discuten las reglas de inferencia

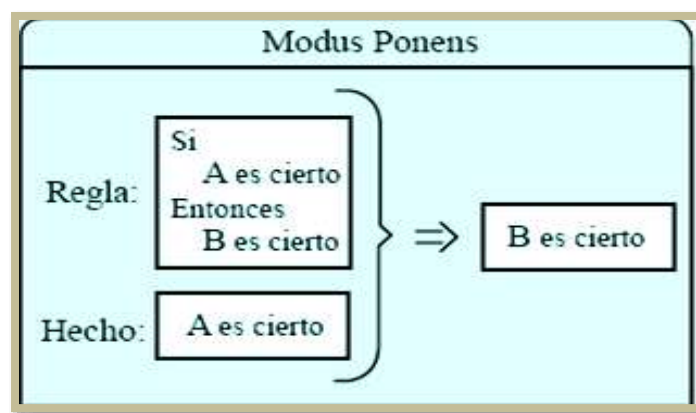
- Modus Ponens,
- Modus Tollens,
- Resolución, y las estrategias de inferencia
- Encadenamiento de reglas,
- Encadenamiento de reglas orientado a un objetivo,
- Compilación de reglas, que son utilizadas por el motor de inferencia para obtener conclusiones simples y compuestas.



Las dos primeras reglas de inferencia se usan para obtener conclusiones simples y el resto de reglas y estrategias para obtener conclusiones compuestas. Nótese, sin embargo, que ninguna de las estrategias anteriores, si se implementan solas, conduce a todas las conclusiones posibles. Por ello, deben implementarse varias reglas y estrategias en el sistema experto para que el motor de inferencia sea capaz de obtener tantas conclusiones como sea posible.

### MODUS PONENS Y MODUS TOLLENS

El *Modus Ponens* es quizás la regla de inferencia más comúnmente utilizada. Se utiliza para obtener conclusiones simples. En ella, se examina la premisa de la regla, y si es cierta, la conclusión pasa a formar parte del conocimiento. Como ilustración, supóngase que se tiene la regla, “Si *A* es cierto, entonces *B* es cierto” y que se sabe además que “*A* es cierto.” Entonces, tal como muestra la Figura 1.7, la regla *Modus Ponens* concluye que “*B* es cierto.” Esta regla de inferencia, que parece trivial, debido a su familiaridad, es la base de un gran número de sistemas expertos.



**FIGURA 1.7.** Una ilustración de la regla de inferencia *Modus Ponens*.

La regla de inferencia *Modus Tollens* se utiliza también para obtener conclusiones simples. En este caso se examina la conclusión y si es falsa, se concluye que la premisa también es falsa. Por ejemplo, supóngase de nuevo que se tiene la regla, “Si *A* es cierto, entonces *B* es cierto” pero se sabe que “*B* es falso.” Entonces, utilizando la regla *Modus Ponens* no se puede obtener ninguna conclusión, pero, tal como se muestra en la Figura 1.8, la regla *Modus Tollens* concluye que “*A* es falso.” Aunque muy simple y con muchas aplicaciones útiles, la regla *Modus Tollens* es menos utilizada que la *Modus Ponens*. Por ello, la regla *Modus Ponens* se mueve hacia adelante, es decir, de la premisa a la conclusión de una regla, mientras que la regla *Modus Tollens* se mueve hacia atrás, es decir, de la conclusión a la premisa. Las dos reglas de inferencia no deben ser vistas como alternativas sino como complementarias.

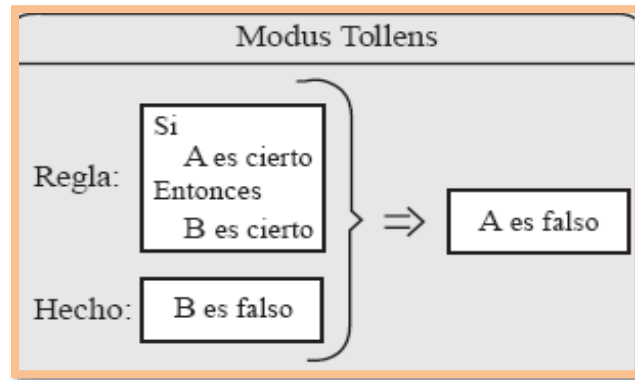


FIGURA 1.8. Una ilustración de la regla Modus Tollens.

La regla Modus Ponens necesita información de los objetos de la premisa para concluir, mientras que la regla Modus Tollens necesita información sobre los objetos de la conclusión. De hecho, para un motor de inferencia que solamente utiliza Modus Ponens, la incorporación de la regla de inferencia Modus Tollens puede ser considerada como una expansión de la base de conocimiento mediante la adición de reglas, tal como ilustra el ejemplo que sigue. Ejemplo 2.3 La regla Modus Tollens equivale a una expansión de la base de conocimiento. Supóngase que la base de conocimiento consiste sólo en la Regla 1, que se muestra en la Figura 1.9. Se puede utilizar la regla de inferencia Modus Tollens para “invertir” la Regla 1 y obtener alguna conclusión cuando se tiene información sobre los objetos de su conclusión.

$\bar{B}$  , entonces  $\bar{A}$  .” En este caso de Regla 1, utilizando la equivalencia

$$A = C \text{ y } B = C \Leftrightarrow A = F \text{ o } B = F,$$

Entonces, aplicar la regla Modus Tollens a la regla “Si A, entonces B” es equivalente a aplicar la regla Modus Ponens a la regla “Si se obtiene la Regla 1b, que se muestra en la Figura 1.10. Por ello, utilizar ambas, las reglas Modus Ponens y Modus Tollens cuando la base de conocimiento contiene sólo la Regla 1, es equivalente a usar la regla Modus Ponens cuando la base de conocimiento contiene ambas, la Regla 1 y la Regla 1b. Por otra parte, el rendimiento del motor de inferencia depende del conjunto de reglas en su base de conocimiento. Hay situaciones en las que el motor de inferencia puede concluir utilizando un conjunto de reglas, pero no puede, utilizando otro (aunque éstos sean lógicamente equivalentes). A continuación se da un ejemplo ilustrativo.



# *Control de la Coherencia*

## TEMA 4



### Competencia:

**Implementar un control de coherencias en sistemas expertos.**





## Tema 04: Control de la Coherencia

En situaciones complejas, incluso verdaderos expertos pueden dar información inconsistente (por ejemplo, reglas inconsistentes y/o combinaciones de hechos no factibles). Por ello, es muy importante controlar la coherencia del conocimiento tanto durante la construcción de la base de conocimiento como durante los procesos de adquisición de datos y razonamiento. Si la base de conocimiento contiene información inconsistente (por ejemplo, reglas y/o hechos), es muy probable que el sistema experto se comporte de forma poco satisfactoria y obtenga conclusiones absurdas.



### El objetivo del control de la coherencia consiste en:

1. Ayudar al usuario a no dar hechos inconsistentes, por ejemplo, dándole al usuario las restricciones que debe satisfacer la información demandada.
2. Evitar que entre en la base de conocimiento cualquier tipo de conocimiento inconsistente o contradictorio.

El control de la coherencia debe hacerse controlando la coherencia de las reglas y la de los hechos.

Objetos		Conclusio		Conclusiones contradictorias
<i>A</i>	<i>B</i>	Re	Regla	
<i>C</i>	<i>C</i>	<i>B</i>	<i>B</i> =	Sí
<i>C</i>	<i>F</i>	=	<i>F B</i>	Sí
<i>F</i>	<i>C</i>		= <i>F</i>	No
<i>F</i>	<i>F</i>	<i>C</i>	–	No

TABLA 2.8. Una tabla de verdad que muestra que las Reglas 1 y 2 son coherentes.

## COHERENCIA DE REGLAS

Reglas coherentes. Un conjunto de reglas se denomina coherente si existe, al menos, un conjunto de valores de todos los objetos que producen conclusiones no contradictorias. En consecuencia, un conjunto coherente de reglas no tiene porque producir conclusiones no contradictorias para todos los posibles conjuntos de valores de los objetos. Es decir, es suficiente que exista un conjunto de valores que conduzcan a conclusiones no contradictorias.



Ejemplo 2.13 Conjunto de reglas incoherentes. Considérense las cuatro reglas siguientes, que relacionan dos objetos A y B binarios {C, F }:

- Regla 1: Si  $A = C$ , entonces  $B = C$ .
- Regla 2: Si  $A = C$ , entonces  $B = F$ .
- Regla 3: Si  $A = F$ , entonces  $B = C$ .
- Regla 4: Si  $A = F$ , entonces  $B = F$ .

**Entonces, pueden obtenerse las siguientes conclusiones:**

1. Las Reglas 1–2 son coherentes puesto que, tal como se muestra en la Tabla 2.8, para  $A = F$ , no producen conclusiones.
2. Las Reglas 1–3 son coherentes puesto que para  $A = F$  y  $B = C$ , producen una conclusión ( $B = C$ ) (véase la Tabla 2.9).
3. Las Reglas 1–4 son incoherentes porque producen conclusiones contradictorias para todos los posibles valores de A y B, tal como se ve en la Tabla 2.10.



Nótese que un conjunto de reglas puede ser coherente, aunque algunos conjuntos de valores puedan producir conclusiones inconsistentes. Estos conjuntos de valores se llaman valores no factibles. Por ejemplo, las Reglas 1–2 son coherentes, aunque producen conclusiones inconsistentes en todos los casos en que

$A = C$ . en consecuencia el subsistema de control de coherencia eliminara automáticamente el valor C de la lista de posibles valores del objeto A, permitiendo de esta forma al usuario seleccionar solo valores factibles de los objetos.

Objetos		Conclusiones			Conclusiones contradictorias
A	B	Regla 1	Regla 2	Regla 3	
C	C	$B = C$	$B = F$	–	Sí
C	F	$B = C$	$B = F$	–	Sí
F	C	–	–	$B =$	No
F	F	–	–	C	Sí
				$B =$	
				C	

TABLA 2.9. Una tabla de verdad que muestra que las Reglas 1–3 son coherentes.

Objetos		Conclusiones				Conclusiones contradictorias
A	B	Regla 1	Regla 2	Regla 3	Regla 4	
C	C	$B =$	$B = F$	–	–	Sí Sí
		$C B$	$B = F$	–	–	Sí Sí
C	F	$= C$	–	$B =$	$B = F$	
		–	–	C	$B = F$	
F	C	–		$B =$		
				C		
F	F					
F						

TABLA 2.10. Una tabla de verdad que muestra que las Reglas 1-4 son incoherentes.

**Definición de Valores no factibles:** Se dice que un valor  $a$  para el objeto  $A$  no es factible si las conclusiones obtenidas al hacer  $A = a$  contradicen cualquier combinación de valores del resto de los objetos.

Por ello, cualquier valor no factible debe ser eliminado de la lista de valores posibles de su correspondiente objeto para eliminar la posibilidad de que el motor de inferencia pueda obtener conclusiones inconsistentes.

1. Las dos primeras reglas implican que  $A = C$ , puesto que  $A = C$  siempre conduce a conclusiones inconsistentes. Por tanto, el valor  $A = C$  deber ser eliminado automáticamente de la lista de valores factibles de  $A$ . Dado que  $A$  es binario, entonces resulta  $A = F$  (el único valor posible).

2. Las tres primeras reglas implican que  $A = F$  y  $B = C$ . Por tanto, el valor  $B = F$  deberá valores factibles de  $B$ .

3. Las primeras cuatro reglas implican que  $A = C$ ,  $A = F$ ,  $B = C$  y  $B = F$ . Por tanto, los valores  $\{C, F\}$  son eliminados de las listas de valores de  $A$  y  $B$ , con lo que las listas de valores factibles de todos los objetos están vacías, lo que implica que las cuatro reglas son incoherentes.



Nótese que es suficiente realizar la comprobación de la coherencia de las reglas sólo una vez, tras ser introducida cada regla, y que todos los valores no factibles pueden ser eliminados de sus correspondientes listas, nada más ser detectados. El conjunto de reglas que forman el conocimiento debe ser coherente; en otro caso, el sistema podrá obtener conclusiones erróneas. Por ello, antes de añadir una regla a la base de conocimiento, hay que comprobar la consistencia de esta regla con el resto de ellas, incluidas en la base de conocimiento. Si la regla fuese consistente con el resto de reglas, se añadiría a la base de conocimiento; en caso contrario, se devolvería al experto humano para su corrección.



Ejemplo de Coherencia de reglas. Supóngase que se tienen los cuatro objetos:  $A \in \{0, 1\}$ ,  $B \in \{0, 1\}$ ,  $C \in \{0, 1, 2\}$  y  $D \in \{0, 1\}$ . Considérense las cuatro reglas:

- Regla 1: Si  $A = 0$  y  $B = 0$ , entonces  $C = 0$ .
- Regla 2: Si  $A = 0$  y  $D = 0$ , entonces  $C = 1$ .
- Regla 3: Si  $A = 0$  y  $B = 0$ , entonces  $C = 1$ .
- Regla 4: Si  $A = 0$ , entonces  $B = 0$ .
- Regla 5: Si  $B = 0$ , entonces  $A = 1$ .

Supóngase ahora que se desea añadir las tres últimas reglas a una base de conocimiento que contiene las dos primeras reglas. Entonces, las Reglas 1 y 3 son inconsistentes, puesto que tienen la misma premisa pero diferentes conclusiones.



Por tanto, la Regla 3 debe ser rechazada y el experto humano informado de la razón del rechazo. El experto humano corregirá la regla en cuestión y/o las reglas existentes si fueran incorrectas. La Regla 4 entrará en la base de conocimiento, puesto que es consistente con las Reglas 1 y 2. La Regla 5 es inconsistente con la Regla 4. Por ello, la consistencia de ambas reglas debe ser comprobada antes de pasar a formar parte de la base de conocimiento.

### COHERENCIA DE HECHOS

Los datos o evidencias suministrados por los usuarios deben ser también consistentes en sí y con el conjunto de reglas de la base de datos. Por ello, el sistema no debe aceptar hechos que contradigan el conjunto de reglas y/o el conjunto de hechos existente en cada instante del proceso. Por ejemplo, con una base de conocimiento que contenga las dos primeras reglas del Ejemplo 2.15, el sistema no debe aceptar el conjunto de hechos  $A = 0$ ,  $B = 0$  y  $C = 1$  puesto que contradicen la Regla 1. El sistema debe también comprobar si existe o no, una solución factible e informar al usuario en consecuencia. Si en el ejemplo anterior se trata de dar la información  $A = 0$ ,  $B = 0$  y  $D = 0$ , el sistema debe detectar que no existe ningún valor de  $C$  que sea consistente con la base de conocimiento. Nótese que antes de conocer los valores de los objetos, existe una solución factible. Por ejemplo,  $A = 0$ ,  $B = 0$ ,  $C = 0$  y  $D = 1$  (estos hechos no contradicen la base de conocimiento). Por ello, la inconsistencia surge que los hechos y las reglas sean inconsistentes.

**La coherencia de los hechos puede lograrse mediante las estrategias siguientes:**

1. Eliminar todos los valores no factibles (los que contradicen el conjunto de reglas y/o hechos) de los objetos una vez detectados. Cuando se pregunte al usuario por información sobre los valores de un conjunto de objetos, el sistema experto debería aceptar sólo los valores de cada objeto que sean consistentes con las reglas y con el conocimiento previo. Considerase, por ejemplo, la base de conocimiento del Ejemplo y supóngase que al sistema experto se le ha dado la información  $A = 0$  y  $C = 1$ ; entonces el sistema debe saber que  $B = 0$ . Por ello, este valor debe ser eliminado de la lista de posibles valores del objeto B.

2. El motor de inferencia debe comprobar que los hechos conocidos no contradicen el conjunto de reglas. En la situación anterior, por ejemplo, el sistema no debe aceptar el conjunto de hechos  $A = 1$ ,  $B = 1$  y  $C = 2$ . Si el sistema no elimina los valores no factibles, entonces el usuario podrá dar evidencias contradictorias tales como  $\text{Pago} = \text{autorizado}$  y  $\text{NIP} = \text{incorrecto}$  en el Ejemplo 2.1 Por ello, tan pronto como se dé la primera evidencia,  $\text{Pago} = \text{autorizado}$ , el sistema debe seleccionar sólo los valores del NIP que no conduzcan a conclusiones contradictorias.

3. Suministrar al usuario una lista de objetos a los que no se ha asignado valores previamente. Para cada uno de los objetos, mostrar y aceptar sólo sus valores factibles. Actualizar continuamente la base de conocimiento, es decir, tan pronto como se dé un hecho o se obtenga una conclusión, y eliminar los valores no factibles. El motor de inferencia obtiene todas las conclusiones posibles examinando, y posiblemente concluyendo, las reglas tan pronto como una simple unidad de información llega al sistema.



# Lecturas Recomendadas

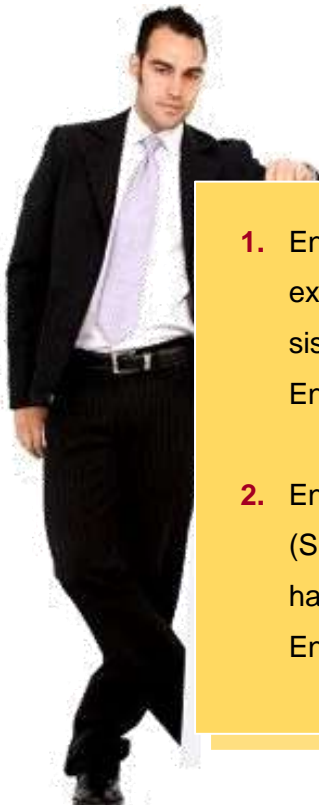
## ❖ APUNTES DE SISTEMAS EXPERTOS

<http://www.dccia.ua.es/dccia/inf/assignaturas/FIA/apuntesse.pdf>

## ❖ SISTEMAS EXPERTOS EN LA TOMA DE DECISIONES

<http://cdigital.uv.mx/bitstream/123456789/28498/1/Drouaillet%20Pumarino.pdf>

# Actividades y Ejercicios



1. En un documento en Word realice un cuadro comparativo y un sistema experto. Además mencione cinco ejemplos de aplicaciones concretas de sistemas expertos. Indicando el hardware y software usado.  
Envíalo a través de **"Sistemas Expertos"**.
2. En un documento en Word describa la arquitectura básica de un sistema (S.) experto (E.), proponga tres ejemplos con aplicaciones, donde se haga evidente la arquitectura usada.  
Envíalo a través de **"Arquitectura de S. E."**.

# Autoevaluación

- 1) **Una definición de sistema experto seria:**
  - a. El sistema Hardware.
  - b. El sistema Software.
  - c. Memorizar información.
  - d. Maquinas que piensan y razonan.
  - e. Procesar información.
  
- 2) **Se utilizan sistemas expertos que operan automáticamente los semáforos y regulan el flujo del tráfico en las calles de una ciudad y en los ferrocarriles, este es un ejemplo de:**
  - a. Control de tráfico.
  - b. Transacciones bancarias.
  - c. Problemas de planificación.
  - d. Diagnóstico médico.
  - e. Problemas de tráfico.
  
- 3) **Las etapas en el desarrollo de un sistema experto son:**
  - a. 5
  - b. 2
  - c. 3
  - d. 7
  - e. 8
  
- 4) **Un conjunto de reglas se denomina \_\_\_\_\_ si existe, al menos, un conjunto de valores de todos los objetos que producen conclusiones no contradictorias.**
  - a. Inteligente.
  - b. Asociadas.
  - c. Controlado.
  - d. Satisfecho.
  - e. Coherente.
  
- 5) **La robótica es una de las áreas que utiliza la:**
  - a. Almacenamiento Interno.
  - b. Inteligencia Artificial.
  - c. Inteligencia Almacenada.
  - d. Actuación inteligente.
  - e. Ideas Asociadas.

- 6) Uno de los objetivos del control de la coherencia consiste en**
- a. No permitir igualdad de valores.
  - b. No permitir programas de software.
  - c. Permitir todo tipo de información y conocimiento aun cuando este sea contradictorio.
  - d. Evitar que entre en la base de conocimiento cualquier conocimiento inconsistente o contradictorio.
  - e. Probar que los sistemas no cometan errores.
- 7) Señala cuál de las siguientes alternativas menciona a las etapas en el desarrollo de un sistema experto.**
- a. Planteamiento de problema, herramienta de desarrollo, probar prototipo.
  - b. Ejecución de órdenes, probar prototipo, usuarios.
  - c. Planteamiento de problema, adquisición de conocimientos, base de datos.
  - d. Probar prototipo, people soft, construir prototipo.
  - e. Interface de usuario, planificación, ejecución de órdenes.
- 8) Es un subsistema controla la consistencia de la base de datos y evita que unidades de conocimiento inconsistentes entren en la misma.**
- a. Control de coherencia.
  - b. Sistema de datos.
  - c. Control de consistencia de los procesos.
  - d. Motor de inferencia.
  - e. Interface de usuarios.
- 9) ¿Cuáles son los dos elementos importantes que intervienen en los sistemas basados en reglas?**
- a. Elemento dinámico y aplicaciones.
  - b. Base de datos y base de conocimientos.
  - c. Base de conocimientos y los datos.
  - d. Memoria de trabajo y datos.
  - e. Conjunto de objetos y conjunto de reglas.
- 10) Es una afirmación lógica que relaciona dos o más objetos e incluye dos partes, la premisa y la conclusión:**
- a. Conclusión.
  - b. Regla.
  - c. Operadores.
  - d. Objeto-valor.
  - e. Expresión lógica.



## UNIDAD DE APRENDIZAJE I: SISTEMAS EXPERTOS BASADOS EN REGLAS

Los sistemas expertos son máquinas que piensan y razonan como un experto lo haría en una cierta especialidad o campo. Por ejemplo, un sistema experto en diagnóstico médico requerirá como datos los síntomas del paciente, los resultados de análisis clínicos y otros hechos relevantes, y, utilizando éstos, buscaría en una base de datos la información necesaria para poder identificar la correspondiente enfermedad.

Los problemas con los que pueden tratar los sistemas expertos pueden clasificarse en dos tipos: problemas esencialmente deterministas y problemas esencialmente estocásticos. Consecuentemente, los sistemas expertos pueden clasificarse en dos tipos principales según la naturaleza de problemas para los que están diseñados: deterministas y estocásticos.

Los sistemas basados en reglas son una herramienta eficiente para tratar estos problemas. Las reglas deterministas constituyen la más sencilla de las metodologías utilizadas en sistemas expertos. La base de conocimiento contiene el conjunto de reglas que definen el problema, y el motor de inferencia saca las conclusiones aplicando la lógica clásica a estas reglas.

En situaciones complejas, incluso verdaderos expertos pueden dar información inconsistente (por ejemplo, reglas inconsistentes y/o combinaciones de hechos no factibles). Por ello, es muy importante controlar la coherencia del conocimiento tanto durante la construcción de la base de conocimiento como durante los procesos de adquisición de datos y razonamiento. Si la base de conocimiento contiene información inconsistente (por ejemplo, reglas y/o hechos), es muy probable que el sistema experto.

## UNIDAD 2



# Sistemas Expertos Basados en Probabilidad

# Introducción

## **a) Presentación y contextualización**

Los temas que se tratan en la presente Unidad Temática, tienen por finalidad que el estudiante conozca los Sistemas expertos probabilísticos que han demostrado resolver muchos problemas que se crían necesitaban ciertas habilidades que sólo se encuentran en los seres humanos (por ejemplo, la habilidad de pensar, observar, memorizar, aprender, ver, oler, etc.). Sin embargo, el trabajo realizado en las tres últimas décadas por investigadores procedentes de varios campos, muestra que muchos de estos problemas pueden ser formulados y resueltos por máquinas.

## **b) Competencia**

**Comprende los sistemas expertos basados en probabilidades relacionados con el cálculo.**

## **c) Capacidades**

1. Interpretar la construcción de modelos probabilísticos.
2. Describir la base de conocimientos de un sistema experto.
3. Reconoce los grafos como una técnica para representar procesos.
4. Emplea grafos definidos en procesos de sistemas expertos.

## **d) Actitudes**

- ✓ Presenta iniciativa en la investigación para profundizar los sistemas expertos.
- ✓ Promueve el desarrollo de ejercicios prácticos basados en probabilidades.

## **e) Presentación de Ideas básicas y contenido esenciales de la Unidad:**

**La Unidad de Aprendizaje 02: Sistemas Expertos Basados en Probabilidad,** comprende el desarrollo de los siguientes temas:

**TEMA 01: Conceptos Básicos de Probabilidad.**

**TEMA 02: La Base del Conocimiento.**

**TEMA 03: Algunos conceptos sobre grafos.**

**TEMA 04: Tipos de Grafos Dirigidos.**

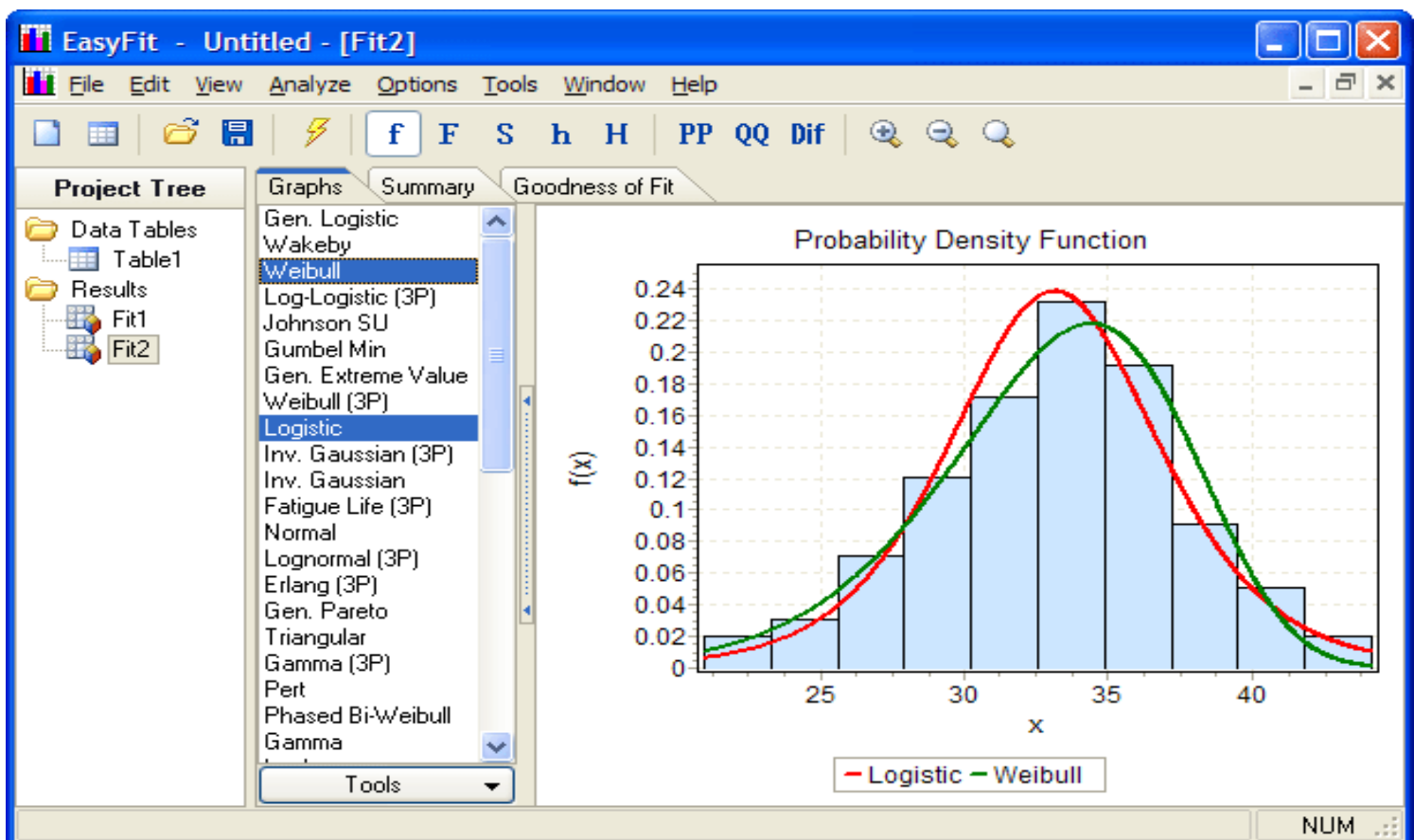
# Conceptos Básicos de Probabilidad

## TEMA 1



Competencia:

Interpretar la construcción de modelos probabilísticos.



# Desarrollo de los Temas



## Tema 01: Conceptos Básicos de Probabilidad

### ALGUNOS CONCEPTOS BÁSICOS DE LA TEORÍA DE LA PROBABILIDAD

En esta sección se introduce el siguiente material básico que será utilizado posteriormente:

- Medida de probabilidad.
- Distribuciones de probabilidad.
- Dependencia e independencia.
- Teorema de Bayes.
- Tipos de errores.



### Medida de probabilidad



Para medir la incertidumbre se parte de un marco de discernimiento dado  $S$ , en el que se incluyen todos los posibles resultados de un cierto experimento como conjunto exhaustivo y mutuamente exclusivo. El conjunto  $S$  se conoce como espacio muestral. Una vez definido este conjunto, el objetivo consiste en asignar a todo subconjunto de  $S$  un número real que mida el grado de incertidumbre sobre su realización. Para obtener medidas con significado físico claro y práctico, se imponen ciertas condiciones o propiedades intuitivas adicionales que definen una clase de medidas que se conocen como medidas de probabilidad.

Una función  $p$  que proyecta los subconjuntos  $A \subseteq S$  en el intervalo  $[0, 1]$  se llama medida de probabilidad si satisface los siguientes axiomas:

- ❖ Axioma 1 (Normalización):  $p(S) = 1$ .
- ❖ Axioma 2 (Actividad): Para cualquier sucesión infinita,  $A_1, A_2, \dots$ , de subconjuntos disjuntos de  $S$ , se cumple la igualdad





El Axioma 1 establece que, independientemente de nuestro grado de certeza, ocurrirá un elemento del conjunto universal  $S$  (es decir, el conjunto  $S$  es exhaustivo).

El Axioma 2 es una fórmula de agregación que se usa para calcular la probabilidad de la unión de subconjuntos disjuntos. Establece que la incertidumbre de un cierto subconjunto es la suma de las incertidumbres de sus partes (disjuntas).



Nótese que esta propiedad también se cumple para sucesiones finitas. De los axiomas anteriores pueden deducirse propiedades muy interesantes de la probabilidad. Por ejemplo:

- Propiedad 1 (Normalización):  $p(\varphi) = 0$ .
- Propiedad 2 (Monotonicidad): Si  $A \subseteq B \subseteq S$ , entonces  $p(A) \leq p(B)$ .
- Propiedad 3 (Continuidad-Consistencia): Para toda sucesión creciente  $A_1 \subseteq A_2 \subseteq \dots$  o decreciente  $A_1 \supseteq A_2 \supseteq \dots$  de subconjuntos de  $S$  se tiene

$$\lim_{i \rightarrow \infty} p(A_i) = p(\lim_{i \rightarrow \infty} A_i).$$

- Propiedad 4 (Inclusión-Exclusión): Dado cualquier par de subconjuntos  $A$  y  $B$  de  $S$ , se cumple siempre la siguiente igualdad:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B).$$



La Propiedad 1 establece que la evidencia asociada a una ausencia completa de información es cero.

La Propiedad 2 muestra que la evidencia de la pertenencia de un elemento a un conjunto debe ser al menos la evidencia de cualquiera de sus subconjuntos. En otras palabras, la evidencia de que un elemento pertenezca a un conjunto dado  $A$  no debe decrecer con la adición de elementos a  $A$ .

La Propiedad 3 puede ser considerada como una propiedad de consistencia o continuidad. Si se eligen dos sucesiones de conjuntos que convergen al mismo subconjunto de  $S$ , se debe obtener la misma evidencia o incertidumbre. La Propiedad 4 establece que las probabilidades de los conjuntos,  $A$ ,  $B$ ,  $A \cap B$ , y  $A \cup B$  no son independientes, sino que están relacionadas. Un ejemplo clásico que ilustra estos axiomas es el del lanzamiento de un dado no trucado.

Aquí el espacio muestral es  $S = \{1, 2, 3, 4, 5, 6\}$ , es decir, el conjunto de los posibles resultados del lanzamiento. Sea  $p(A)$  la probabilidad de que ocurra el suceso  $A$ .

Entonces, por ejemplo, se tiene:

$$p(S) = 1, p(\{1\}) = 1/6, p(\{3\}) = 1/6, \text{ y } p(\{1, 3\}) = p(\{1\}) + p(\{3\}) = 1/3.$$

## Distribuciones de probabilidad

Sea  $\{X_1, \dots, X_n\}$  un conjunto de variables aleatorias discretas y  $\{x_1, \dots, x_n\}$  el conjunto de sus posibles realizaciones. Nótese que las variables aleatorias se denotan con mayúsculas y que sus realizaciones se denotan con minúsculas. Por ejemplo, si  $X_i$  es una variable binaria, entonces  $x_i$  puede ser 1 o 0. Los resultados que siguen son también válidos si las variables son continuas, pero en este caso los símbolos de suma deben sustituirse por integrales.



Sea  $p(x_1, \dots, x_n)$  la función de probabilidad conjunta<sup>1</sup> de las variables de  $X$ , es decir,

$$p(x_1, \dots, x_n) = p(X_1 = x_1, \dots, X_n = x_n)$$


Entonces, la función de probabilidad marginal de la  $i$ -ésima variable se obtiene mediante la fórmula

$$p(x_i) = p(X_i = x_i) = \sum_{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n} p(x_1, \dots, x_n)$$

$$p(x_1, \dots, x_n).$$

El conocimiento de la ocurrencia de un suceso puede modificar las probabilidades de otros sucesos. Por ejemplo, la probabilidad de que un paciente tenga una enfermedad dada puede cambiar tras el conocimiento de los resultados de un análisis de sangre.

Por ello, cada vez que se dispone de nueva información, las probabilidades de los sucesos pueden, y suelen, cambiar. Esto conduce al concepto de probabilidad condicional.



$x$	$y$	$z$	$p(x, y, z)$
0	0	0	0.12
0	0	1	0.18
0	1	0	0.04
0	1	1	0.16
1	0	0	0.09
1	0	1	0.21
1	1	0	0.02
1	1	1	0.18

**Tabla 3.2. Función de probabilidad conjunta de tres variables binarias.**

Probabilidad condicional. Sean  $X$  e  $Y$  dos conjuntos disjuntos de variables tales que  $p(y) > 0$ . Entonces, la probabilidad condicional (función de probabilidad condicionada) de  $X$  dado  $Y = y$  viene dada por

$$p(X = x|Y = y) = p(x|y) = \frac{p(x, y)}{p(y)} \quad \dots(3.5)$$

La ecuación (3.5) implica que la función de probabilidad conjunta de  $X$  e  $Y$  puede escribirse como:

$$p(x, y) = p(y)p(x|y) \quad \dots(3.6)$$

### Dependencia e independencia

Independencia de dos variables. Sean  $X$  e  $Y$  dos subconjuntos disjuntos el conjunto de variables aleatorias  $\{X, \dots, X_n\}$ . Entonces se dice que  $X$  es independiente de  $Y$  si y solamente si:

$$p(x|y) = p(x), \dots\dots\dots (3.7)$$

Para todos los valores posibles  $x$  e  $y$  de  $X$  e  $Y$ ; en otro caso,  $X$  se dice dependiente de  $Y$ .

Nótese que si  $x$  e  $y$  son valores posibles de  $X$  e  $Y$ , entonces  $p(x) > 0$  y  $p(y) > 0$ . Por ello, la condición  $p(y) > 0$  es natural en el sentido de que no puede observarse  $Y = y$  si no se satisface la condición.

$$p(x|y) = p(x) \dots (3.8)$$

La ecuación (3.8) significa que si  $X$  es independiente de  $Y$ , entonces nuestro conocimiento de  $Y$  no afecta nuestro conocimiento sobre  $X$ , es decir,  $Y$  no tiene información sobre  $X$ . También, si  $X$  es independiente de  $Y$ , pueden combinarse (3.6) y (3.8) para obtener  $p(x, y)/p(y) = p(x)$ , que implica

$$p(x, y) = p(x)p(y) \dots (3.9)$$

La ecuación (3.9) indica que si  $X$  es independiente de  $Y$ , entonces la función de probabilidad conjunta de  $X$  e  $Y$  es igual al producto de sus marginales. En realidad, (3.9) es una definición de independencia equivalente a la (3.8). Una propiedad importante de la relación de independencia es su simetría, es decir, si  $X$  es independiente de  $Y$ , entonces  $Y$  es independiente de  $X$ .

**Esto ocurre porque**

$$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{p(x)p(y)}{p(x)} = p(y) \dots (3.10)$$

Por la propiedad de simetría se dice que  $X$  e  $Y$  son independientes o mutuamente independientes. La implicación práctica de la simetría es que si el conocimiento de  $Y$  es relevante (irrelevante) para  $X$ , entonces el conocimiento de  $X$  es relevante (irrelevante) para  $Y$ .

Los conceptos de dependencia e independencia de dos variables aleatorias pueden ser extendidos al caso de más de dos variables aleatorias como sigue:

Independencia de un conjunto de variables: Las variables aleatorias  $\{X_1, \dots, X_m\}$  se dice que son independientes si y sólo si

$$p(x_1, \dots, x_m) = \prod_{i=1}^m p(x_i), \quad \dots(3.11)$$

Para todos los valores posibles  $x_1, \dots, x_m$  de  $X_1, \dots, X_m$ . En otro caso, se dice que son dependientes. En otras palabras,  $\{X_1, \dots, X_m\}$  se dicen independientes si y sólo si su función de probabilidad conjunta es igual al producto de sus funciones de probabilidad marginal. Nótese que (3.11) es una generalización de (3.9).

Nótese también que si  $X_1, \dots, X_m$  son condicionalmente independientes dado otro subconjunto  $Y_1, \dots, Y_n$ , entonces

$$p(x_1, \dots, x_m | y_1, \dots, y_n) = \prod_{i=1}^m p(x_i | y_1, \dots, y_n). \quad \dots\dots\dots(3.12)$$

Una implicación importante de la independencia es que no es rentable obtener información sobre variables independientes, pues es irrelevante. Es decir, independencia significa irrelevancia.

Dependencia e independencia condicional. Sean  $X$ ,  $Y$  y  $Z$  tres conjuntos disjuntos de variables, entonces  $X$  se dice condicionalmente independiente de  $Y$  dado  $Z$ , si y sólo si;

$$p(x | z, y) = p(x | z)$$

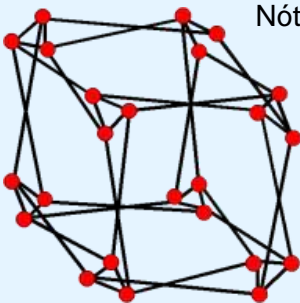


**Para todos los valores posibles de  $x$ ,  $y$  y  $z$  de  $X$ ,  $Y$  y  $Z$ ; En otro caso  $X$  e  $Y$  se dicen condicionalmente dependientes dado  $Z$ .**

Cuando  $X$  e  $Y$  son condicionalmente independientes dado  $Z$ , se escribe  $I(X, Y | Z)$ . La relación  $I(X, Y | Z)$  se denomina relación de independencia condicional. Similarmente, cuando  $X$  e  $Y$  son condicionalmente dependientes dado  $Z$ , se escribe  $D(X, Y | Z)$  que se conoce como una relación de dependencia condicional. A veces escribimos  $I(X, Y | Z)_P$  o  $D(X, Y | Z)_P$  para indicar que la relación se deriva, o es implicada, por el modelo probabilístico asociado a la probabilidad  $p$  (la función de probabilidad conjunta).

La definición de independencia condicional lleva en sí la idea de que una vez que es conocida  $Z$ , el conocimiento de  $Y$  no altera la probabilidad de  $X$ . En otras palabras, si  $Z$  ya es conocida, el conocimiento de  $Y$  no añade información alguna sobre  $X$ . Una definición alternativa, pero equivalente, de independencia condicional es

$$p(x, y | z) = p(x | z) p(y | z).$$



Nótese que la independencia (incondicional) puede ser tratada como un caso particular de la independencia condicional. Por ejemplo, se puede escribir  $I(X, Y | \phi)$  para indicar que  $X$  e  $Y$  son incondicionalmente independientes, donde  $\phi$  es el conjunto vacío. Nótese, sin embargo, que  $X$  e  $Y$  pueden ser independientes incondicionalmente pero condicionalmente dependientes dado  $Z$ , es decir, la relación de independencia condicional  $I(X, Y | \phi)$  y la de dependencia condicional  $D(X, Y | Z)$  pueden satisfacerse simultáneamente.

Por ejemplo, para determinar si  $X$  e  $Y$  son independientes, se necesita comprobar si  $p(x, y) = p(x) p(y)$  para todos los valores posibles de  $x$  e  $y$ .

También se puede determinar si cualesquiera dos variables son condicionalmente independientes dada una tercera variable.



Por ejemplo, para comprobar si  $X$  e  $Y$  son condicionalmente independientes dado  $Z$ , es necesario comprobar si  $p(x | y, z) = p(x, y, z)/p(y, z) = p(x | z)$  para todos los valores posibles de  $x$ ,  $y$  y  $z$ . Para ello, se calculan las probabilidades cuyos valores se muestran en la Tabla 3.6. En esta tabla puede verse que  $p(x | y, z) = p(x | z)$  y, por tanto,  $D(X, Y | Z)$ . Por ello, la función de probabilidad conjunta de la Tabla 3.2 implica que  $X$  e  $Y$  son incondicionalmente independientes,  $I(X, Y | \phi)$ , aunque son condicionalmente dependientes dado  $Z$ ,  $D(X, Y | Z)$ .

$$p(x|y, z) = \frac{p(x, y, z)}{p(y, z)},$$

$$p(x|z) = \frac{p(x, z)}{p(z)},$$

y	z	x	$p(x y, z)$
0	0	0	$12/21 \approx 0.571$
0	0	1	$9/21 \approx 0.429$
0	1	0	$18/39 \approx 0.462$
1	0	0	$4/6 \approx 0.667$
1	0	1	$2/6 \approx 0.333$
1	1	0	$16/34 \approx 0.471$



**TABLA 3.6. Funciones de probabilidad obtenida de la función de probabilidad conjunta de la Tabla 3.2.**

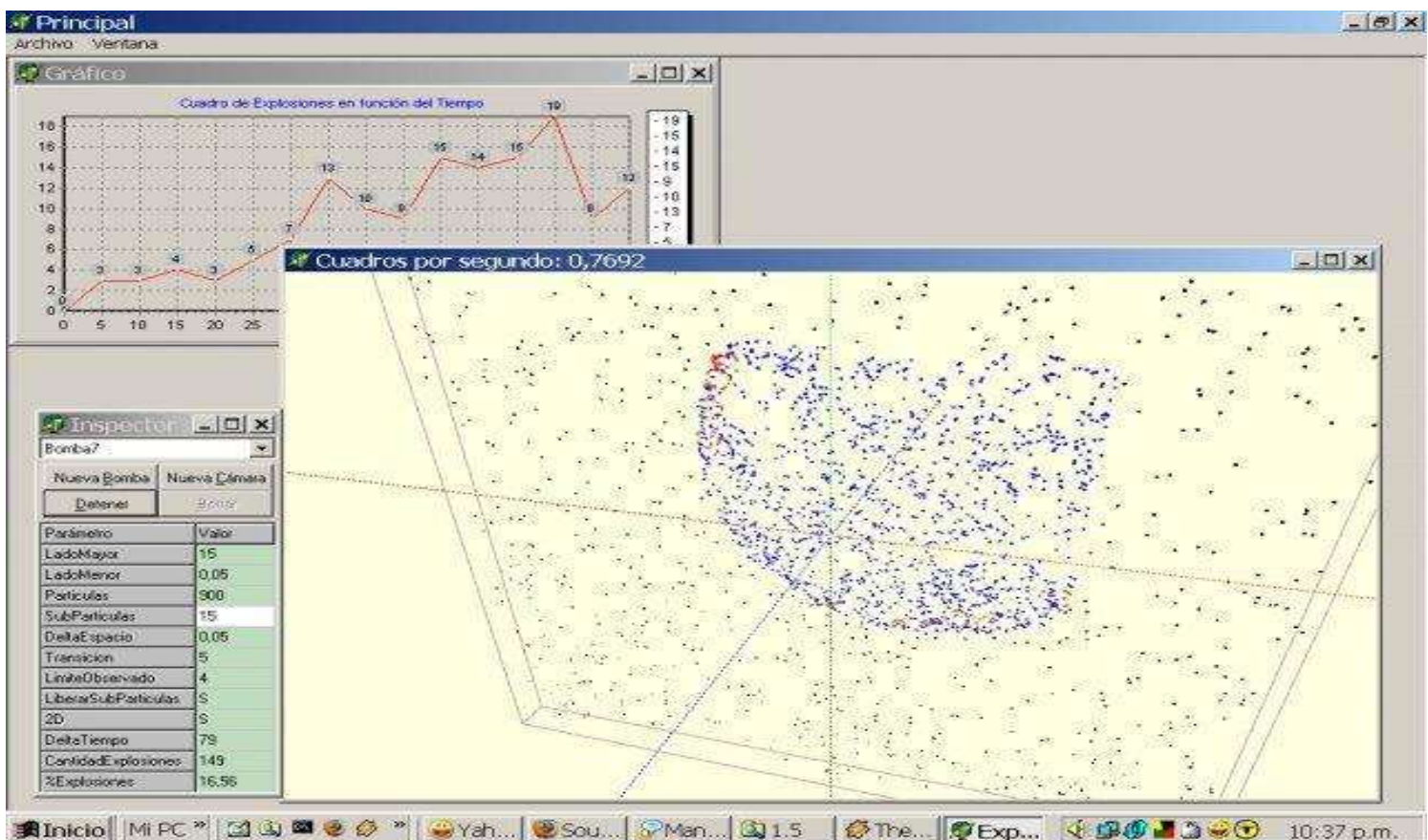
# *La Base del Conocimiento*

## TEMA 2



### Competencia:

**Describir la base de conocimientos de un sistema experto.**





## Tema 02: La Base del Conocimiento



La base de conocimiento de un sistema experto probabilístico consiste en un conjunto de variables,  $\{X_1, \dots, X_n\}$ , y una función de probabilidad conjunta definida sobre ellas,  $p(x_1, \dots, x_n)$ . Por ello, para construir la base de conocimiento de un sistema experto probabilístico, se necesita definir la función de probabilidad conjunta de las variables.

El modelo más general posible se basa en especificar directamente la función de probabilidad conjunta; es decir, asignar un valor numérico (parámetro) a cada una de las posibles combinaciones de valores de las variables. Desgraciadamente, la especificación directa de la función de probabilidad conjunta implica un gran número de parámetros. Por ejemplo, con  $n$  variables binarias, la función de probabilidad conjunta más general tiene  $2^n$  parámetros (las probabilidades  $p(x_1, \dots, x_n)$  para toda posible realización  $\{x_1, \dots, x_n\}$  de las variables, un número tan grande que no hay ordenador en el mundo capaz de almacenarlo incluso para un valor de  $n$  tan pequeño como 50.



Esta fue una de las primeras críticas al uso de la probabilidad en los sistemas expertos. Sin embargo, en la mayor parte de las situaciones prácticas, muchos subconjuntos de variables pueden ser independientes o condicionalmente independientes. En tales casos, se pueden obtener simplificaciones del modelo más general teniendo en cuenta la estructura de independencia de las variables. Esto suele dar lugar a una reducción importante del número de parámetros.

**En esta sección se discuten los siguientes ejemplos de tales simplificaciones:**

1. El Modelo de Síntomas Dependientes (MSD).
2. El Modelo de Síntomas Independientes (MSI).
3. El Modelo de Síntomas Relevantes Independientes (MSRI).
4. El Modelo de Síntomas Relevantes Dependientes (MSRD).

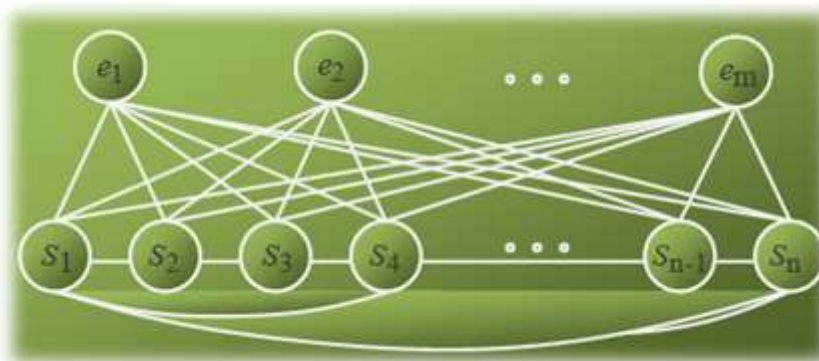
Sin embargo, estos cuatro modelos son modelos ad hoc que se aplican principalmente en el campo médico. Modelos probabilísticos más generales y potentes (por ejemplo, modelos de redes de Markov, modelos de redes Bayesianas, y modelos especificados condicionalmente).

## EL MODELO DE SÍNTOMAS DEPENDIENTES

En este modelo, se supone que los síntomas son dependientes pero que las enfermedades son independientes entre sí, dados los síntomas. Todo síntoma se conecta con los demás síntomas y con todo valor posible de E (indicando dependencia).

Entonces la función de probabilidad conjunta para el MSD puede escribirse como:

$$p(e_i, s_1, \dots, s_n) = p(s_1, \dots, s_n)p(e_i | s_1, \dots, s_n).$$



Una ilustración gráfica del modelo de síntomas dependientes.



### El modelo de síntomas independientes

Debido a la imposibilidad de trabajar con el modelo anterior en muchos casos prácticos, resulta necesario proceder a la simplificación del modelo. Una simplificación posible consiste en suponer que, para una enfermedad dada, los síntomas son condicionalmente independientes entre sí. El modelo resultante se denomina modelo de síntomas independientes (MSI). El MSI se ilustra en la Figura 3.7, donde los síntomas no están ligados, para indicar la independencia.

Puesto que los síntomas se suponen condicionalmente independientes dada la enfermedad, se tiene;

$$p(s_1, \dots, s_n | e_i) = \prod_{j=1}^n p(s_j | e_i).$$

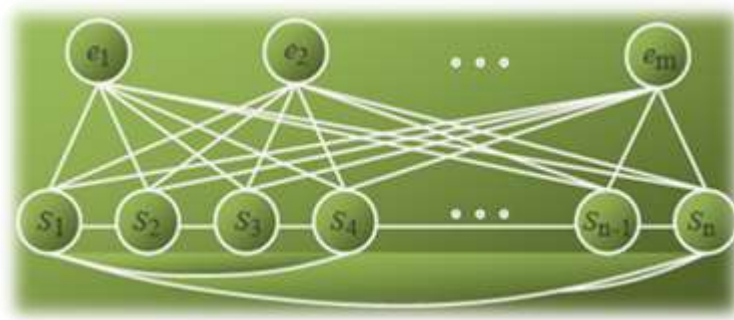


FIGURA 3.7. Una ilustración gráfica del modelo de síntomas independientes.

Por ello, se puede escribir la función de probabilidad conjunta de la enfermedad E dados los síntomas  $s_1, \dots, s_n$  como

$$\begin{aligned} p(e_i | s_1, \dots, s_n) &= \frac{p(e_i) p(s_1, \dots, s_n | e_i)}{p(s_1, \dots, s_n)} \\ &= \frac{p(e_i) \prod_{j=1}^n p(s_j | e_i)}{p(s_1, \dots, s_n)} \end{aligned} \quad (3.26)$$

$$\propto p(e_i) \prod_{j=1}^n p(s_j | e_i). \quad (3.27)$$

La ecuación muestra como la hipótesis de independencia modifica las probabilidades de todas las enfermedades cuando se conocen nuevos síntomas. Por ello, la probabilidad inicial de la enfermedad  $e_i$  es  $p(e_i)$ , pero tras conocer los síntomas  $s_j$ , para  $j = 1, \dots, k$ , resulta proporcional a  $p(s_j | e_i)$ . Nótese que cada nuevo síntoma conduce a un nuevo factor. Nótese también que  $p(s_1, \dots, s_n)$ , en el denominador es una constante de normalización que no es necesario calcular directamente.

- ❖ Las probabilidades marginales  $p(e_i)$ , para todos los valores posibles de la enfermedad  $E$ .
- ❖ Las probabilidades condicionales  $p(s_j | e_i)$ , para todos los valores posibles del síntoma  $S_j$  y la enfermedad  $E$ .



Por ello, con las hipótesis de independencia de los síntomas, el número de parámetros se reduce considerablemente. Con  $m$  enfermedades posibles y  $n$  síntomas binarios, el número total de parámetros es  $m(n + 1) - 1$ . Por ejemplo, con  $m = 100$  enfermedades y  $n = 200$  síntomas, se tienen 20,099 parámetros en el MSI en vez de más de 1062 parámetros para el MSD. Ejemplo de El Modelo de síntomas independientes. Para ilustrar el MSI, se utilizan los historiales clínicos de dos centros médicos, cada uno de ellos consta de  $N = 1000$  pacientes; dos valores de la enfermedad ( $g$  y  $g^-$ ); y tres síntomas,  $D$ ,  $V$  y  $P$ .

Centro Médico 1						
		$g$		$\bar{g}$		Total
		$d$	$\bar{d}$	$d$	$\bar{d}$	
$v$	$p$	220	95	4	31	350
	$\bar{p}$	25	10	5	50	90
$\bar{v}$	$p$	220	95	9	76	400
	$\bar{p}$	25	10	12	113	160
Total		490	210	30	270	1000

Centro Médico 2						
		$g$		$\bar{g}$		Total
		$d$	$\bar{d}$	$d$	$\bar{d}$	
$v$	$p$	140	210	0	0	350
	$\bar{p}$	0	0	30	60	90
$\bar{v}$	$p$	280	0	0	120	400
	$\bar{p}$	70	0	0	90	160
Total		490	210	30	270	1000

**TABLA 3.13. Números de pacientes clasificados por una enfermedad  $G$  y tres síntomas,  $D$ ,  $V$  y  $P$  en dos centros médicos.**

$e$	$p(e)$	$e$	$d$	$p(d e)$	$e$	$v$	$p(v e)$	$e$	$p$	$p(p e)$
$\bar{g}$	0.3	$\bar{g}$	0	0.9	$\bar{g}$	0	0.7	$\bar{g}$	0	0.6
$g$	0.7	$\bar{g}$	1	0.1	$\bar{g}$	1	0.3	$\bar{g}$	1	0.4
		$g$	0	0.3	$g$	0	0.5	$g$	0	0.1
		$g$	1	0.7	$g$	1	0.5	$g$	1	0.9

**TABLA 3.14. Probabilidades requeridas para la especificación del MSI.**

Los datos se resumen en la Tabla 3.13. Nótese que los datos del Centro Médico 1 son los mismos que los de la Figura 3.1, pero dados ahora en forma tabular, en vez de forma gráfica. Para especificar el MSI, se necesita la probabilidad marginal,  $p(e)$ , de la enfermedad y las probabilidades condicionales de cada síntoma dada cada enfermedad,  $p(d|e)$ ,  $p(v|e)$  y  $p(p|e)$ . Estas probabilidades se extraen de la Tabla 3.13 y se dan en la Tabla 3.14. Nótese que sólo 7 parámetros son libres. Un aspecto interesante de los dos conjuntos de datos es que aunque son muy diferentes, conducen a idénticas probabilidades, como se muestra en la Tabla 3.14.

En la Tabla 3.15 se da la probabilidad condicional de  $E$  dadas varias combinaciones de los síntomas para los dos centros médicos. Nótese que;

			Centro Médico 1		Centro Médico 2	
$d$	$v$	$p$	Valor Real	MSI	Valor Real	MSI
0	0	0	0.08	0.08	0.00	0.08
0	0	1	0.56	0.56	0.00	0.56
0	1	0	0.17	0.18	0.00	0.18
0	1	1	0.75	0.74	1.00	0.74
1	0	0	0.68	0.66	1.00	0.66
1	0	1	0.96	0.96	1.00	0.96
1	1	0	0.83	0.82	0.00	0.82
1	1	1	0.98	0.98	1.00	0.98

**TABLA 3.15. La probabilidad condicional  $p(g | d, v, p)$**

$$p(\bar{g}|d, v, p) = 1 - p(g|d, v, p)$$

Los valores exactos se calculan directamente de la Tabla 3.13 utilizando la definición de probabilidad condicional dada en (3.3). Los valores de las columnas etiquetadas MSI se calculan aplicando la fórmula para el MSI en (3.27). Por ejemplo, para el Centro Medico 1, el valor de  $p(g|d, v, p)$  se calcula mediante

$$p(g|d, v, p) = \frac{p(g, d, v, p)}{p(d, v, p)} = \frac{220}{220 + 4} = 0.98.$$

El valor de  $p(g|d, v, p)$  según el MSI se calcula usando

$$p(g|d, v, p) \propto p(g)p(d|g)p(v|g)p(p|g) = 0.7 \times 0.7 \times 0.5 \times 0.9 = 0.2205,$$

$$p(\bar{g}|d, v, p) \propto p(\bar{g})p(d|\bar{g})p(v|\bar{g})p(p|\bar{g}) = 0.3 \times 0.1 \times 0.3 \times 0.4 = 0.0036.$$

Dividiendo 0.2205 por la constante de normalización  $0.2205 + 0.0036 = 0.2241$ , se obtiene  $p(g|d, v, p) = 0.2205/0.2241 = 0.98$  y  $p(\bar{g}|d, v, p) = 0.0036/0.2241 = 0.02$ .



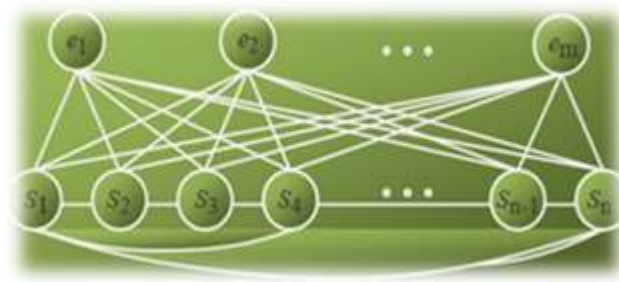
Una comparación entre las probabilidades verdaderas y las correspondientes al MSI de la Tabla 3.15 muestra que los dos conjuntos de probabilidades son parecidos para el Centro Medico 1, pero discrepan notablemente para el Centro Medico 2. Por ejemplo, para el Centro Medico 2 el valor real de  $p(g | d, v, p^-)$  es 0, mientras que el correspondiente al MSI es 0.82.

Esto es una prueba de que el MSI falla al tratar de describir la probabilidad de los datos del Centro Medico 2. Nótese que se tienen dos conjuntos de datos con las mismas probabilidades “a priori” y las mismas verosimilitudes; sin embargo, el MSI es apropiado para reproducir uno de ellos y no, para el otro.

**De este ejemplo puede concluirse que las probabilidades “a priori” y las verosimilitudes no son suficientes para especificar un modelo probabilístico. Por tanto, debe ponerse especial cuidado en la selección del modelo probabilístico a utilizar en un caso dado.**

### Modelo de síntomas relevantes independientes

Se puede conseguir una reducción aún mayor del número de parámetros suponiendo que cada enfermedad tiene un número reducido de síntomas relevantes. En consecuencia, para cada valor  $e$  de la enfermedad  $E$  se seleccionan algunos síntomas relevantes  $S_1, \dots, S_r$  (relativamente pocos frente al total de síntomas) y los restantes síntomas se suponen independientes para ese valor de  $E$ . El MSRI se ilustra en la Figura 3.16. Nótese que para  $e_1$ , el conjunto de síntomas relevantes es  $\{S_1, S_2\}$ ; para  $e_2$ , el conjunto de síntomas relevantes es  $\{S_2, S_3, S_4\}$ ; y así sucesivamente.



**FIGURA 3.16. Una ilustración gráfica del modelo de síntomas Relevantes independientes.**

Por simplicidad de notación, supóngase que  $S_1, \dots, S_{r_i}$  son relevantes para la enfermedad  $e_i$  y que los restantes síntomas  $S_{r_i+1}, \dots, S_n$  son irrelevantes. Según el MSRI,  $p(s_j | e_i)$  se supone idéntica para todos los síntomas que son irrelevantes para la enfermedad  $e_i$ . Entonces la función de probabilidad conjunta de la enfermedad  $e_i$  dados los síntomas  $s_1, \dots, s_n$  puede escribirse como sigue donde  $p_j = p(s_j | e_i)$ , que es la misma para todas las enfermedades para la que  $S_j$  es irrelevante, se obtiene el MSRI. Se deduce que es necesario almacenar las probabilidades siguientes en la base de conocimiento del MSRI:



$$\begin{aligned} p(e_i | s_1, \dots, s_n) &= \frac{p(e_i)p(s_1, \dots, s_n | e_i)}{p(s_1, \dots, s_n)} \\ &= \frac{p(e_i) \prod_{j=1}^{r_i} p(s_j | e_i) \prod_{j=r_i+1}^n p(s_j | e_i)}{p(s_1, \dots, s_n)} \\ &= \frac{p(e_i) \prod_{j=1}^{r_i} p(s_j | e_i) \prod_{j=r_i+1}^n p_j}{p(s_1, \dots, s_n)} \end{aligned} \quad (3.28)$$



- ✓ Las probabilidades marginales  $p(e_i)$ , para todos los valores posibles de la enfermedad  $E$ .
- ✓ Las probabilidades condicionales  $p(s_j | e_i)$ , para cada valor posible de  $E$  y cada uno de sus correspondientes síntomas relevantes.
- ✓ Las probabilidades  $p_j$ , para cada valor posible de  $E$  que tiene al menos un síntoma irrelevante. (Esto implica que  $p_j = p(s_j | e_i)$  es idéntica para todos los síntomas irrelevantes para  $e_i$ ).

La ecuación implica que en la base de conocimiento se necesita almacenar las probabilidades de todos los síntomas relevantes para cada enfermedad, y la misma probabilidad para todos los síntomas irrelevantes para cada valor de  $E$ . Por ello, si se tienen  $m$  posibles enfermedades y  $n$  síntomas binarios, el número de parámetros en el MSRI es;

$$m - 1 + n - a + \sum_{i=1}^m r_i,$$

Donde  $r_i$  es el número de síntomas relevantes para la enfermedad  $e_i$  y  $a$  es el número de síntomas que son relevantes para todas las enfermedades. El número de parámetros se reduce significativamente cuando  $r_i$  es mucho menor que  $n$ . Por ejemplo, con 100 enfermedades y 200 síntomas, si  $r_i = 10$  para todas las enfermedades, el número de parámetros en el MSRI se reduce de 20,099



para el MSI a 1,299 para el MSRI. Nótese que se puede obtener el MSRI a partir del MSI, sin más que imponer algunas restricciones adicionales en los parámetros del MSI, puesto que en el MSRI las probabilidades  $p(s_j | e_i)$  deben ser las mismas para todos los síntomas irrelevantes para las enfermedades  $e_i$ .

El número de restricciones es parámetros en el MSI,  $(m(n + 1) - 1)$ , menos el número de restricciones.

$$a - n + \sum_{j=1}^n n_j,$$

En total, se tiene

$$m(n + 1) - 1 + n - a - \sum_{j=1}^n n_j,$$

# Algunos Conceptos sobre Grafos

## TEMA 3



### Competencia:

Reconocer los grafos como una técnica para representar procesos.



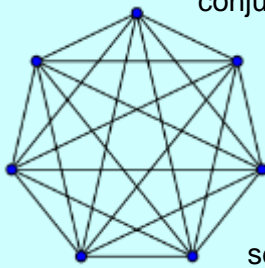


## Tema 03: Algunos Conceptos sobre Grafos

### CONCEPTOS BÁSICOS Y DEFINICIONES

Supóngase un conjunto de objetos  $X = \{X_1, X_2, \dots, X_n\}$  que pueden relacionarse entre sí. El conjunto  $X$  puede ser representado gráficamente por una colección de nodos o vértices, asociando un nodo a cada elemento de  $X$ . Estos nodos pueden conectarse por aristas, indicando las relaciones existentes entre los mismos. Una arista entre los nodos  $X_i$  y  $X_j$  se denotará mediante  $L_{ij}$ .

Así mismo, el conjunto de todas las aristas se denotará por  $L = \{L_{ij} \mid X_i \text{ y } X_j \text{ están conectados}\}$ . Por tanto, un grafo puede definirse de forma intuitiva mediante el conjunto de nodos,  $X$ , y las relaciones entre los mismos,  $L$ . En el siguiente ejemplo se ilustra esta idea intuitiva. A continuación se introduce una definición formal.



Ejemplo de Grafos. La Figura 4.1 es un ejemplo de un grafo compuesto de seis nodos  $X = \{A, B, \dots, G\}$  y de un conjunto de seis aristas,

$L = \{L_{AB}, L_{AC}, L_{BD}, L_{CE}, L_{DF}, L_{DG}\}$

Los nodos están representados por círculos y las aristas por líneas que unen los nodos correspondientes.

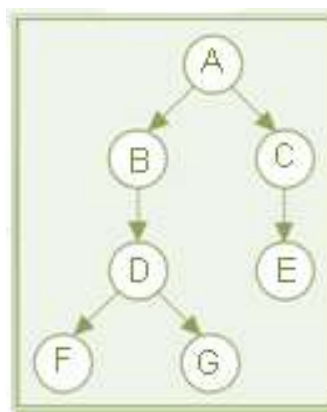


FIGURA 4.1. Ejemplo de un grafo o red.

El concepto de grafo puede definirse de forma más general. Por ejemplo, puede permitirse que dos nodos estén conectados por más de una arista, o incluso que un nodo esté conectado consigo mismo. Sin embargo, en el campo de los sistemas expertos, los grafos se utilizan para representar un conjunto de variables proposicionales (nodos), y unas relaciones de dependencia entre ellas (aristas). Por tanto, no es necesario que dos nodos estén unidos por más de una arista, o que una arista una un nodo consigo mismo.

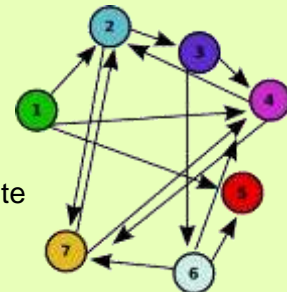
Las aristas de un grafo pueden ser dirigidas o no dirigidas, dependiendo de si se considera o no, el orden de los nodos. En la práctica, esta distinción dependerá de la importancia del orden en que se relacionen los objetos.

### Arista dirigida:

Dado un grafo  $G = (X, L)$ , si  $L_{ij} \in L$  y  $L_{ji} \notin L$ , la arista  $L_{ij}$  entre los nodos  $X_i$  y  $X_j$  se denomina dirigida y se denota mediante  $X_i \rightarrow X_j$ .

### Arista no dirigida:

Dado un grafo  $G = (X, L)$ , si  $L_{ij} \in L$  y  $L_{ji} \in L$ , la arista  $L_{ij}$  se denomina no dirigida y se denota mediante  $X_i - X_j$  o  $X_j - X_i$ .



### Grafo dirigido y no dirigido:

Un grafo en el cual todas las aristas son dirigidas se denomina grafo dirigido, y un grafo en el que todas sus aristas son no dirigidas se denomina no dirigido.

Por tanto, en un grafo dirigido es importante el orden del par de nodos que definen cada arista, mientras que en un grafo no dirigido, el orden carece de importancia.

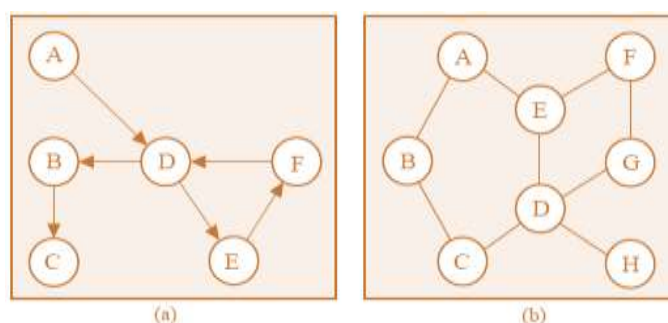


FIGURA 4.2. Ejemplos de un grafo dirigido (a), y uno no dirigido (b).

El grafo de la Figura 4.2(a) esta definido por:

$$X = \{A, B, C, D, E, F\},$$

$$L = \{A \rightarrow D, B \rightarrow C, D \rightarrow B, F \rightarrow D, D \rightarrow E, E \rightarrow F\}$$

Mientras que para el grafo de la Figura 4.2 (b) se tiene

$$X = \{A, B, C, D, E, F, G, H\},$$

$$L = \{A - B, B - C, C - D, D - E, E - A, E - F, F - G, G - D, D - H\}$$

Conjunto adyacente. Dado un grafo  $G = (X, L)$  y un nodo  $X_i$ , el conjunto adyacente del nodo  $X_i$  es el conjunto de nodos que son directamente alcanzables desde  $X_i$ , es decir,  $Ady(X_i) = \{X_j \in X \mid L_{ij} \in L\}$

Esta definición proporciona una descripción alternativa de un grafo mediante un conjunto de nodos,  $X$ , y los conjuntos adyacentes de cada uno de los nodos en  $X$ ; es decir, el grafo  $(X, L)$  puede ser representado de forma equivalente mediante  $(X, Ady)$ , donde  $X = \{X_1, \dots, X_n\}$  es el conjunto de nodos y  $Ady = \{Ady(X_1), \dots, Ady(X_n)\}$  es la lista de conjuntos adyacentes. Como se verá más adelante, esta forma de representación de un grafo es muy conveniente desde un punto de vista computacional.



**Ejemplo 4.3 Conjuntos adyacentes.** El grafo dirigido dado en la Figura 4.2(a) tiene asociados los siguientes conjuntos de nodos adyacentes:

$$Ady(A) = \{D\}, \quad Ady(B) = \{C\}, \quad Ady(C) = \phi,$$

$$Ady(D) = \{B, E\}, \quad Ady(E) = \{F\}, \quad Ady(F) = \{D\}.$$

Por otra parte, los conjuntos adyacentes del grafo no dirigido de la Figura 4.2 (b) son:

$$Ady(A) = \{B, E\}, \quad Ady(B) = \{A, C\},$$

$$Ady(C) = \{B, D\}, \quad Ady(D) = \{C, E, G, H\}$$

$$Ady(E) = \{A, D, F\}, \quad Ady(F) = \{E, G\},$$

$$Ady(G) = \{D, F\}, \quad Ady(H) = \{D\}.$$

Por tanto, los grafos mostrados en la Figura 4.2 pueden ser definidos de forma equivalente por  $(X, L)$  o por  $(X, Ady)$ .



El conjunto adyacente de un nodo  $X_i$  contiene los nodos que son directamente alcanzables desde  $X_i$ . Por tanto, comenzando en un nodo dado y pasando de forma sucesiva a uno de sus nodos adyacentes, se puede formar un camino a través del grafo. Como se verá más adelante, el concepto de camino entre dos nodos juega un papel central en la teoría de grafos.

Camino entre dos nodos. Un camino del nodo  $X_i$  al nodo  $X_j$  es una sucesión de nodos  $(X_{i1}, \dots, X_{ir})$ , comenzando en  $X_{i1} = X_i$  y finalizando en  $X_{ir} = X_j$ , de forma que existe una arista del nodo  $X_{ik}$  al nodo  $X_{ik+1}$ ,  $k = 1, \dots, r - 1$ , es decir,

$$X_{ik+1} \in \text{Ady}(X_{ik}), \quad k = 1, \dots, r - 1$$

La longitud del camino,  $(r - 1)$ , se define como el número de aristas que contiene. En el caso de grafos no dirigidos, un camino  $(X_{i1}, \dots, X_{ir})$  puede representarse mediante  $X_{i1} - \dots - X_{ir}$ , indicando el carácter no dirigido de las aristas. De modo similar, otra forma de representar un camino en un grafo dirigido es mediante

$$X_{i1} \rightarrow \dots \rightarrow X_{ir}.$$

## Ejemplo:



Considérese el grafo dirigido dado en la Figura 4.2(a). Existe un único camino de longitud 2 de D a F en este grafo,  $D \rightarrow E \rightarrow F$ . Por otra parte, existe un camino de A a B de longitud 2,  $A \rightarrow D \rightarrow B$ , y otro de longitud 5,  $A \rightarrow D \rightarrow E \rightarrow F \rightarrow D \rightarrow B$ . Obsérvese que, por el contrario, no existe ningún camino de B a A. Por otra parte, existe al menos un camino entre cada par de nodos del grafo no dirigido de la Figura 4.2 (b). Por ejemplo, algunos de los caminos entre A a H son;

$A - E - D - H$ , de longitud 3,

$A - B - C - D - H$ , de longitud 4, y

$A - E - F - G - D - H$ , de longitud 5.



Nótese que en un grafo dirigido han de tenerse en cuenta las direcciones de las aristas para formar un camino. Por ejemplo, en el grafo dirigido de la Figura 4.2(a) existe un camino de A a C ( $A \rightarrow D \rightarrow B \rightarrow C$ ), pero no existe ningún camino que una los nodos en sentido inverso.

## Camino cerrado:

Un camino  $(X_{i1}, \dots, X_{ir})$  se dice que es cerrado si el nodo inicial coincide con el final, es decir,  $X_{i1} = X_{ir}$ .

Ejemplo

Camino cerrado. El camino  $D \rightarrow G \rightarrow F \rightarrow D$  en el grafo dirigido de la Figura 4.3(a) es un camino cerrado. El grafo no dirigido dado en la Figura 4.3 (b) contiene varios caminos cerrados como, por ejemplo, el camino  $A - B - C - D - E - A$ .



Si un camino contiene un nodo más de una vez, entonces el camino contiene un sub-camino cerrado. Por ejemplo, en el grafo de la Figura 4.3(b), el camino  $C - D - E - F - G - D - H$  contiene dos veces el nodo D. Por tanto, este camino ha de contener un sub-camino cerrado:  $D - E - F - G - D$ . Eliminando este camino cerrado, se puede hallar un camino más corto entre los nodos extremos,  $C - D - H$ .

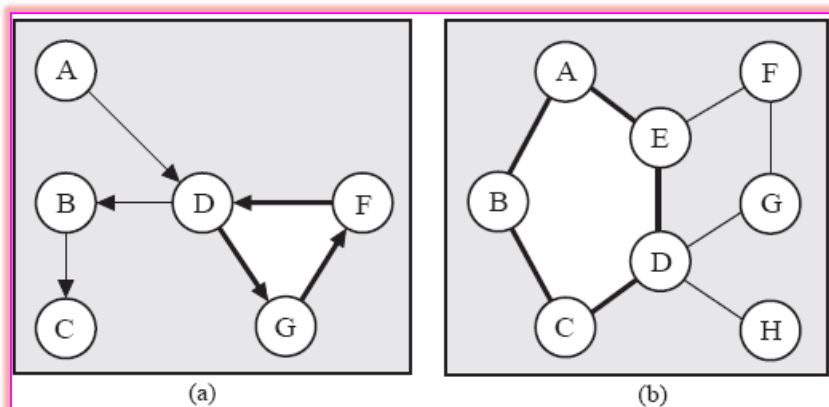


FIGURA 4.3. Ejemplos de caminos cerrados en un grafo dirigido (a) y en un grafo no dirigido (b).

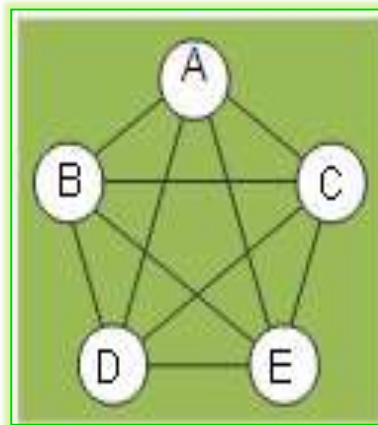


## Características de los grafos no dirigidos

### Definiciones y conceptos básicos

**Grafo completo.** Un grafo no dirigido se denomina completo si contiene una arista entre cada par de nodos.

Por tanto, existe un único grafo completo de  $n$  nodos. Este grafo se denota por  $K_n$ . Por ejemplo, la Figura 4.4 muestra una representación gráfica de  $K_5$ .



**FIGURA 4.4.** Grafo completo de cinco nodos.

**Conjunto completo.** Un subconjunto de nodos  $S$  de un grafo  $G$  se denomina completo si existe una arista en  $G$  para cada par de nodos en  $S$ .

Una consecuencia inmediata de esta definición es que cualquier par de nodos adyacentes en un grafo forma un conjunto completo.

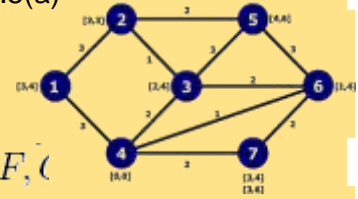
Por ejemplo, el grafo de la Figura 4.3(b) no contiene conjuntos completos con más de dos nodos. Por el contrario, el grafo mostrado en la Figura 4.5(a) contiene dos subconjuntos completos de tres nodos:  $\{D, E, G\}$  y  $\{E, F, G\}$ .

Los conjuntos completos maximales de un grafo desempeñan un papel fundamental en la caracterización de su estructura topológica.

**Conglomerado:** Un conjunto completo de nodos  $C$  se denomina un conglomerado si no es subconjunto propio de otro conjunto completo, es decir, si es maximal.

Ejemplo de Conglomerados. El grafo mostrado en la Figura 4.5(a) Contiene los siguientes conglomerados:

$$C_1 = \{A, B\}, C_2 = \{B, C\}, C_3 = \{C, D\}, C_4 = \{D, H\}, C_5 = \{D, E, G\}, C_6 = \{E, F, G\}$$



Sin embargo, si se añade alguna arista al grafo, alguno de estos conglomerados ya no será un conjunto maximal y el conjunto de conglomerados del nuevo grafo será distinto. Por ejemplo, en el grafo de la Figura 4.5(b), obtenido añadiendo tres aristas al grafo de la Figura 4.5(a), los conjuntos  $C_1$ ,  $C_2$ ,  $C_3$  y  $C_7$  ya no son completos. El nuevo grafo contiene solamente cinco conglomerados:

$$C_1 = \{A, B, D, E\}, C_2 = \{B, C, D\}, C_3 = \{D, H\}, C_4 = \{D, E, G\}, \text{ y } C_5 = \{E, F, G\}.$$

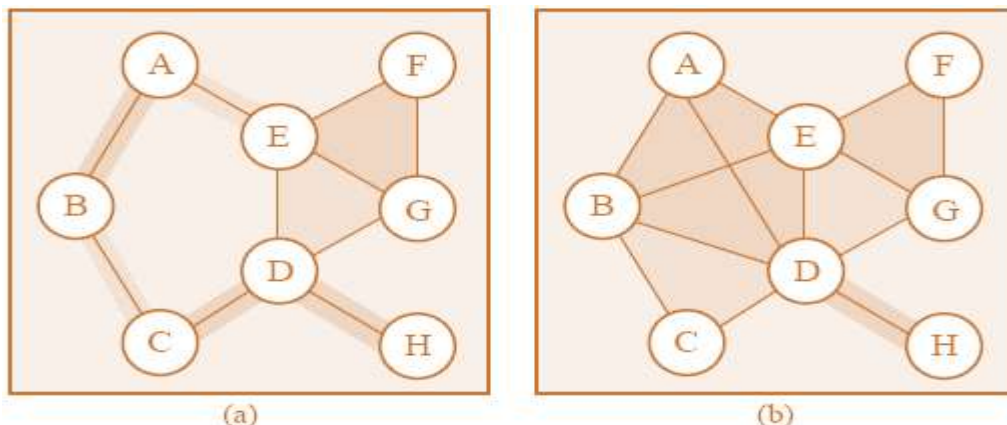


FIGURA 4.5. Ejemplo de los conglomerados asociados a dos grafos distintos.

**Bucle.** Un bucle es un camino cerrado en un grafo no dirigido.

Ejemplo de Bucle: Considérese el grafo no dirigido mostrado en la Figura 4.5 (b). El camino cerrado  $A - B - C - D - E - A$  es un bucle de longitud 5. Obsérvese que si en un bucle se reemplaza un camino entre dos nodos por un camino alternativo, se obtiene un nuevo bucle. Por ejemplo, si se reemplaza la arista  $D-E$  por el camino  $D-G-F-E$  en el bucle anterior, se obtiene un nuevo bucle de longitud 7:  $A - B - C - D - G - F - E - A$ .

Vecinos de un nodo. El conjunto de nodos adyacentes a un nodo  $X_i$  en un grafo no dirigido se denomina conjunto de vecinos de

$$X_i, Vec(X_i) = \{X_j \mid X_j \in Ady(X_i)\}$$

Nótese que en el caso de grafos no dirigidos, el conjunto de nodos adyacentes a un nodo dado coincide con el conjunto de vecinos de dicho nodo. Por ejemplo, los nodos sombreados,  $\{A, D, F\}$ , en la Figura 4.6 son los vecinos del nodo E.

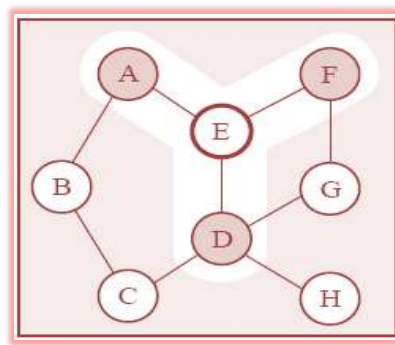


FIGURA 4.6. Conjunto de vecinos del nodo E.

Frontera de un conjunto de nodos: La unión de los conjuntos de vecinos de los nodos de un conjunto dado,  $S$ , excluyendo los nodos de  $S$ , se denomina la frontera de  $S$  y se denota por  $Frn(S)$ .

$$Frn(S) = \left( \bigcup_{X_i \in S} Vec(X_i) \right) \setminus S,$$

Donde  $X \setminus S$  es el conjunto de nodos de  $X$  excluyendo los de  $S$ .



Por ejemplo, los nodos sombreados en la Figura 4.7,  $\{A, C, F, G, H\}$ , son la frontera del conjunto  $\{D, E\}$ . En el caso de que  $S$  contenga un único nodo, la frontera se reduce al conjunto de vecinos.

### Tipos de grafos no dirigidos

En muchas situaciones prácticas es importante conocer si existe un camino entre un par de nodos dados. Por ejemplo, en el campo de los sistemas expertos, los grafos se utilizan para representar relaciones de dependencia entre las variables que componen el sistema. En estos casos, es muy útil conocer el número de posibles caminos entre dos nodos, a efectos de entender la estructura de dependencia contenida en el grafo. Desde este punto de vista, una clasificación útil de los grafos debe tener en cuenta el número de caminos distintos existentes entre cada par de nodos.

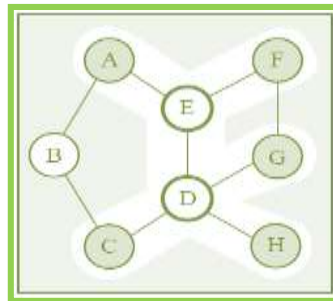


FIGURA 4.7. Frontera del conjunto {D, E}.

Grafos conexos no dirigidos. Un grafo no dirigido se denomina conexo si existe al menos un camino entre cada par de nodos. En caso contrario, el grafo se denomina inconexo. Por ejemplo, el grafo de la Figura 4.7 es un grafo conexo. Sin embargo, el grafo representado en la Figura 4.8 es inconexo pues, por ejemplo, no existe ningún camino entre los nodos A y F. Obsérvese que el grafo mostrado en la Figura 4.8(a) parece conexo a primera vista, pues las aristas se cruzan ocultando este hecho. Esta característica se refleja de forma más directa en la representación gráfica de la Figura 4.8(b).

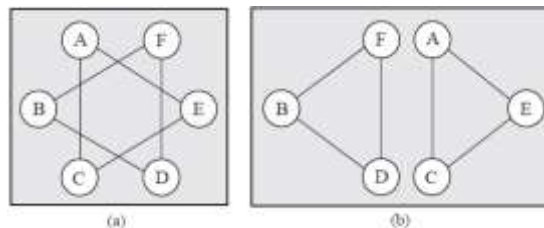


FIGURA 4.8. Dos representaciones distintas del mismo grafo inconexo.

Un grafo inconexo puede dividirse en un conjunto de grafos conexos llamados componentes conexas.

Por ejemplo, el grafo inconexo anterior contiene las componentes conexas  $\{A, C, E\}$  y  $\{B, D, F\}$ . Este hecho hace que, en la práctica, se suponga que los grafos son conexos pues, en caso contrario, podría argumentarse sobre cada una de las componentes conexas del grafo de forma análoga.

La complejidad topológica de un grafo aumenta con el número de caminos distintos entre dos nodos. Por tanto, además de considerar la existencia de un camino entre dos nodos, se ha de considerar también el número de caminos posibles.

**Árbol.** Un grafo conexo no dirigido se denomina un árbol si existe un único camino entre cada par de nodos.

De la definición anterior se deduce que un árbol es un grafo conexo, pero si se elimina una cualquiera de sus aristas, el grafo se vuelve inconexo. De forma similar, se puede deducir que un árbol no contiene bucles, pero si se añade una arista cualquiera al grafo se forma un bucle.





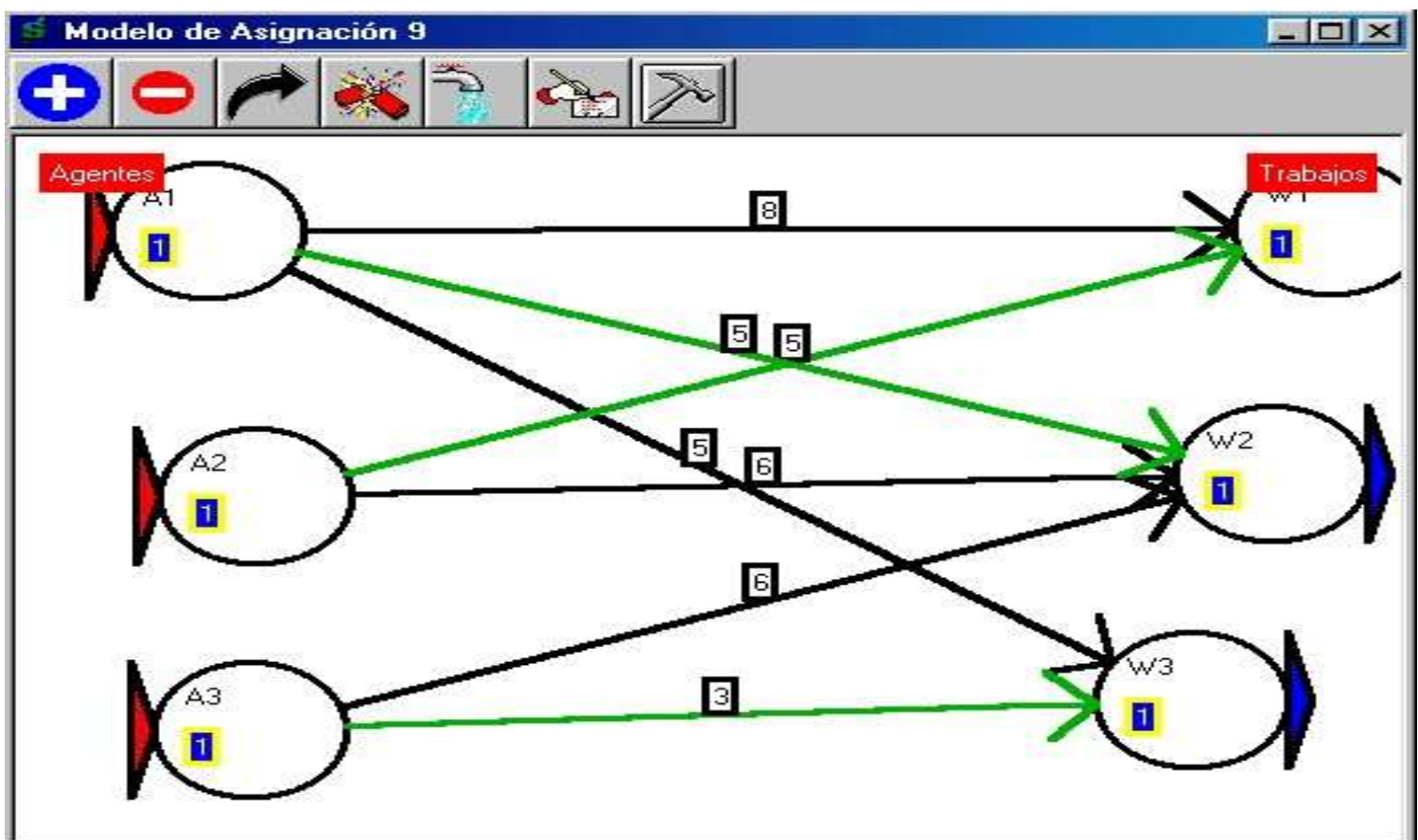
# *Tipos de Grafos Dirigidos*

## TEMA 4



**Competencia:**

**Emplear grafos definidos en procesos de sistemas expertos.**

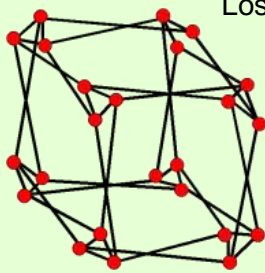




## Tema 04: Tipos de Grafos Dirigidos

**Grafos dirigidos conexos:** Un grafo dirigido se denomina conexo si el grafo no dirigido asociado es conexo; en caso contrario se denomina inconexo.

**Árboles y grafos múltiplemente conexos:** Un grafo dirigido conexo se denomina árbol si el grafo no dirigido asociado es un árbol; en caso contrario se denomina múltiplemente conexo. **Grafos cíclicos y acíclicos:** Un grafo dirigido se denomina cíclico si contiene al menos un ciclo; en caso contrario se denomina grafo dirigido acíclico.



Los grafos dirigidos acíclicos jugarán un papel muy importante más adelante, pues serán la base para construir los modelos probabilísticos conocidos como Redes Bayesianas. Dentro de los grafos dirigidos, los árboles suelen clasificarse en dos tipos, dependiendo del número de aristas que convergen en un mismo nodo. **Grafos simples y poliárboles:** Un árbol dirigido se denomina un árbol simple si cada nodo tiene como máximo un padre; en caso contrario se denomina un poliárbol.

La Figura 4.17 muestra un ejemplo de un árbol simple y un ejemplo de un poliárbol. La Figura 4.18 muestra un grafo cíclico y uno múltiplemente conexo. La Figura 4.19 muestra de modo esquemático estos tipos de grafos dirigidos.

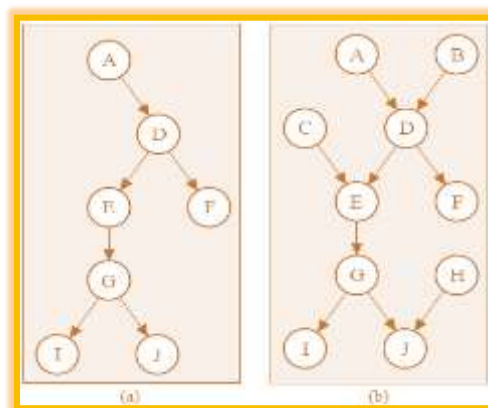


FIGURA 4.17. Ejemplos de grafos dirigidos: árbol simple (a) y poli árbol (b).

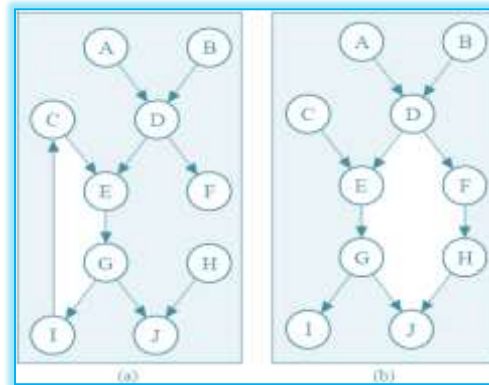


FIGURA 4.18. Ejemplos de grafos dirigidos: grafo cíclico (a) y múltiplemente conexo (b).

### Grafos triangulados

Los grafos triangulados son un tipo especial de grafos no dirigidos que tienen muchas aplicaciones prácticas interesantes en varios campos. Los grafos triangulados también reciben el nombre de circuitos rígidos y grafos cordales. **Cuerda de un bucle:** Una cuerda es una arista que une dos nodos de un bucle y que no pertenece al bucle. Por ejemplo, en el grafo de la Figura 4.20, la arista E - G es una cuerda del bucle E - F - G - D - E. Obsérvese que la cuerda divide el bucle en dos bucles menores: E - F - G - E y E - G - D - E. Por otra parte, el bucle A - B - C - D - E - A no contiene ninguna cuerda.



Dada su estructura, los bucles de longitud 3 son los únicos que no pueden poseer cuerdas. Por ello, estos son los menores elementos en los que puede descomponerse un bucle mediante la incorporación de cuerdas en el grafo.

Los bucles de longitud 3 se denominan triángulos.

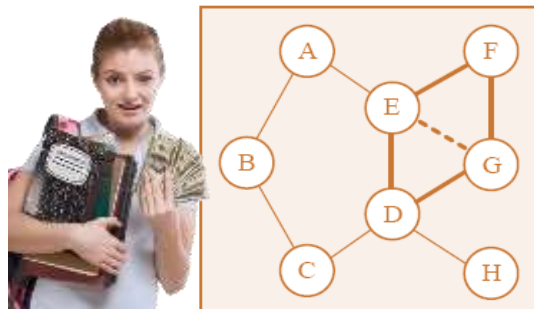


FIGURA 4.20. Ejemplo de un bucle con una cuerda.

**Grafo triangulado:** Un grafo no dirigido se denomina triangulado, o cordal, si cada bucle de longitud mayor o igual que cuatro contiene al menos una cuerda. Ejemplo de un Grafo triangulado. La Figura 4.21(a) muestra un grafo triangulado. El grafo contiene dos bucles de longitud cuatro,  $A-B-E-C-A$  y  $B-C-E-D-B$ , y un bucle de longitud cinco,  $A-B-D-E-C-A$ , y cada uno de ellos tiene al menos una cuerda.

Por otra parte, el grafo de la Figura 4.21(b) no es triangulado, pues contiene al bucle  $A-B-C-D-E-A$ , que no posee ninguna cuerda.

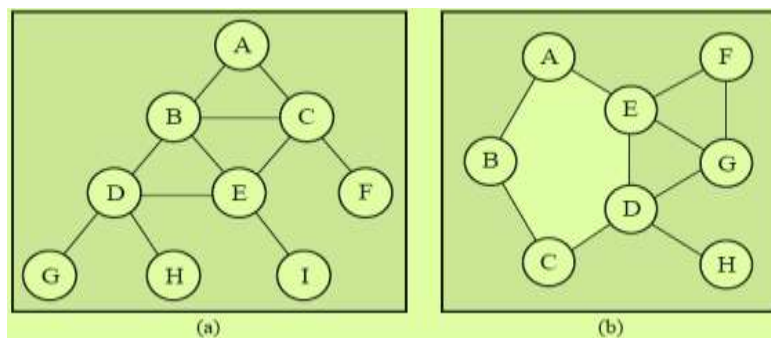


FIGURA 4.21. Ejemplo de grafo triangulado (a) y no triangulado (b).

Si un grafo no es triangulado, es posible convertirlo en triangulado añadiendo cuerdas que dividan los bucles. Este proceso se denomina rellenado o triangulación. Es importante destacar que triangular un grafo no consiste en dividirlo en triángulos. Por ejemplo, el grafo de la Figura 4.21(a) es triangulado y, por tanto, no necesita la adición de aristas extra, como aquellas que se indican mediante líneas de puntos en la Figura 4.22.

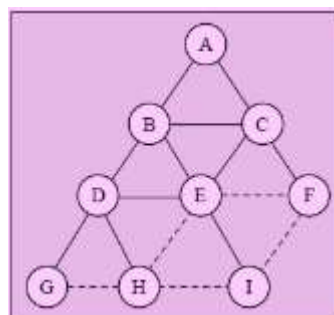
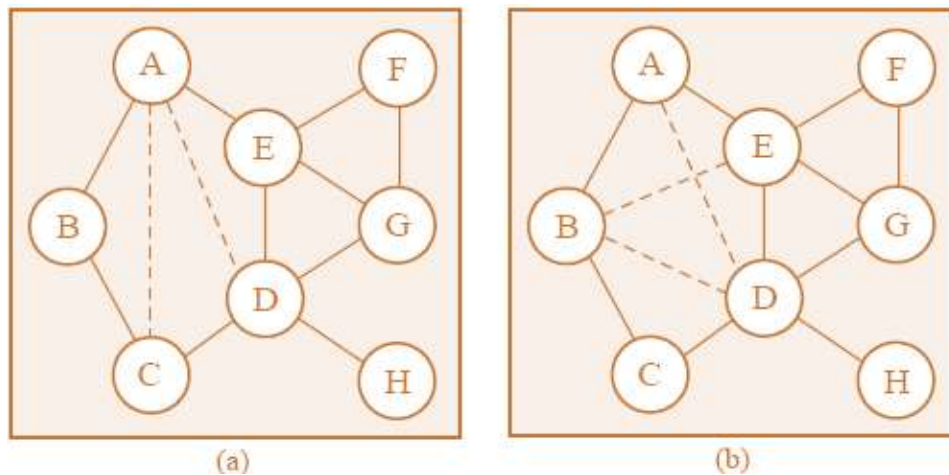


FIGURA 4.22. Triangular no significa dividir en triángulos.

Puesto que un bucle puede romperse de varias formas distintas con una cuerda, existen varias formas distintas de triangular un grafo. Por ejemplo, los dos grafos mostrados en la Figura 4.23 corresponden a dos triangulaciones distintas asociadas con el grafo de la Figura 4.21(b).



**FIGURA 4.23.** Dos triangulaciones distintas del mismo grafo. Las líneas de puntos representan las cuerdas añadidas.



# Lecturas Recomendadas

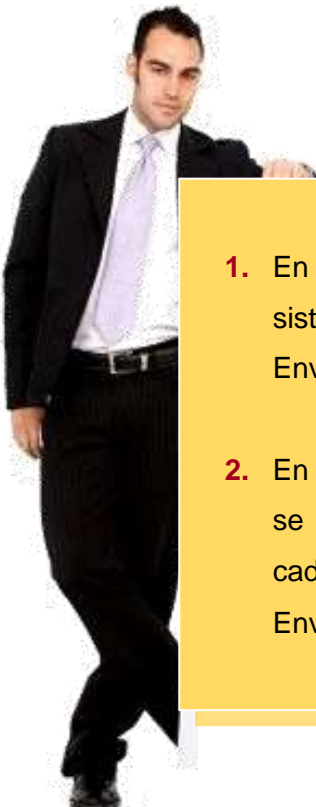
## ❖ ESTRUCTURA DE UN SISTEMA EXPERTO

<https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxqbGNodWN8Z3g6NjZiOGY2ZThjNTJhODJk>

## ❖ DESCUBRIMIENTO DE CONOCIMIENTO BASADO EN GRAFOS

<http://ccc.inaoep.mx/~emorales/Cursos/NvoAprend/grafos.pdf>

# Actividades y Ejercicios



1. En un documento en Word explique la relación entre la probabilidad y los sistemas (S.) expertos (E.). Muestre dos ejemplos concretos.

Envíalo a través de **"Probabilidades en S. E."**.

2. En un documento en Word presente tres ejemplos diferentes sobre cómo se aplica los grafos en sistemas expertos y describa su aplicación en cada uno.

Envíalo a través de **"Grafos"**.



# Autoevaluación

- 1) El conocimiento de la ocurrencia de un suceso puede modificar las - \_\_\_\_\_ de otros sucesos.
  - a. Informaciones.
  - b. Ecuaciones.
  - c. Condiciones.
  - d. Variables.
  - e. Probabilidades.
  
- 2) Una urna tiene ocho bolas rojas, 5 amarilla y siete verdes. Si se extrae una bola al azar calcular la probabilidad de no sacar amarilla.
  - a.  $5/20$
  - b.  $1/20$
  - c.  $2/5$
  - d.  $3/20$
  - e.  $4/5$
  
- 3) Una urna contiene tres bolas rojas y siete blancas. Se extraen dos bolas al azar. Hallar la probabilidad de extraer dos rojas.
  - a.  $9/100$
  - b.  $1/200$
  - c.  $1/3$
  - d.  $3/20$
  - e.  $9/400$
  
- 4) La base de conocimiento de un sistema experto probabilístico consiste en:
  - a. Un conjunto de datos y una función de probabilidad.
  - b. Un conjunto de conocimientos y una función de probabilidad.
  - c. Un conjunto de variables y una función de probabilidad.
  - d. Un conjunto de datos y conocimientos.
  - e. Un conjunto de variables y datos.
  
- 5) En el modelo de síntomas dependientes:
  - a. Los síntomas son independientes.
  - b. El conjunto de datos son independiente.
  - c. El conjunto de variables son dependientes.
  - d. Las enfermedades son independientes.
  - e. La función de probabilidad es dependiente.

- 6) En los grafos los conjuntos pueden ser representados por:
- a. Nodos.
  - b. Graficas.
  - c. Nulos.
  - d. Vacíos.
  - e. Comprensión.
- 7) Una característica de un grafo dirigido es:
- a. Tienen nodos dirigidos.
  - b. Tiene aristas dirigidas.
  - c. Tiene los conjuntos dirigidos.
  - d. Tiene grafos nodos vacíos.
  - e. Su fuerza para entrar en un sistema.
- 8) Se dice que un camino es cerrado
- a. Cuando existen bucles.
  - b. Cuando el nodo final es igual que el nodo inicial.
  - c. Cuando no hay salida.
  - d. Cuando no hay inicio.
  - e. Cuando no hay caminos.
- 9) Los grafos dirigidos acíclicos son la base para construir
- a. Las redes de Ishikawa.
  - b. Aristas Bayesianas.
  - c. Base de datos.
  - d. Base de conocimientos.
  - e. Los modelos probabilísticos.
- 10) Los grafos triangulados también reciben el nombre de:
- a. intersección dinámica.
  - b. Base de conocimientos.
  - c. Circuitos rígidos.
  - d. Redes bayesianas.
  - e. Redes perceptrón.

## UNIDAD DE APRENDIZAJE II: SISTEMAS EXPERTOS BASADOS EN PROBABILIDAD

La probabilidad es un método mediante el cual se obtiene la frecuencia de un suceso determinado mediante la realización de un experimento aleatorio, del que se conocen todos los resultados posibles, bajo condiciones suficientemente estables. La teoría de la probabilidad se usa extensamente en los sistemas expertos.

La base de conocimiento de un sistema experto probabilístico consiste en un conjunto de variables,  $\{X_1, \dots, X_n\}$ , y una función de probabilidad conjunta definida sobre ellas,  $p(x_1, \dots, x_n)$ . Por ello, para construir la base de conocimiento de un sistema experto probabilístico, se necesita definir la función de probabilidad conjunta de las variables.

Supóngase un conjunto de objetos  $X = \{X_1, X_2, \dots, X_n\}$  que pueden relacionarse entre sí. El conjunto  $X$  puede ser representado gráficamente por una colección de nodos o vértices, asociando un nodo a cada elemento de  $X$ . Estos nodos pueden conectarse por aristas, indicando las relaciones existentes entre los mismos. El concepto de grafo puede definirse de forma más general. Por ejemplo, puede permitirse que dos nodos estén conectados por más de una arista, o incluso que un nodo esté conectado consigo mismo.

Un grafo dirigido se denomina conexo si el grafo no dirigido asociado es conexo; en caso contrario se denomina inconexo. Árboles y grafos múltiplemente conexos: Un grafo dirigido conexo se denomina árbol si el grafo no dirigido asociado es un árbol; en caso contrario se denomina múltiplemente conexo. Grafos cíclicos y acíclicos: Un grafo dirigido se denomina cíclico si contiene al menos un ciclo; en caso contrario se denomina grafo dirigido acíclico.

## UNIDAD<sup>3</sup>



# Modelos Probabilísticos y Gráficos

# Introducción

## **a) Presentación y contextualización**

Los temas que se tratan en la presente unidad temática, tienen por finalidad que el estudiante conozca los algoritmos para grafos que necesitan un mecanismo de búsqueda para explorar los nodos y aristas de un grafo. Estos métodos son la base para la construcción de los algoritmos. La exploración de un grafo comienza en un nodo inicial y consiste en la definición de un criterio para moverse hacia adelante y hacia atrás a través de las aristas del grafo, estos se pueden ejecutar siguiendo modelos probabilísticos.

## **b) Competencia**

**Describe los modelos probabilísticos y gráficos en sistemas expertos para el desarrollo de los cálculos.**

## **c) Capacidades**

1. Interpreta la construcción de modelos probabilísticos
2. Describe los modelos definidos gráficamente empleando sus técnicas.
3. Aplica las Técnicas y herramientas formales de análisis de sistemas desde los modelos definidos gráficamente.
4. Emplea las extensiones de los modelos gráficos para desarrollar de manera técnica las probabilidades.

## **d) Actitudes**

- ✓ Muestra autonomía para resolver tus cálculos aplicando modelos probabilísticos.
- ✓ Iniciativa para profundizar y ampliar los conocimientos con respecto a los modelos probabilísticos.

## **e) Presentación de Ideas básicas y contenido esenciales de la Unidad:**

**La Unidad de Aprendizaje 03: Modelos Probabilísticos y Gráficos**, comprende el desarrollo de los siguientes temas:

**TEMA 01:** Construcción de Modelos Probabilísticos.

**TEMA 02:** Modelos Definidos Gráficamente I.

**TEMA 03:** Modelos Definidos Gráficamente II.

**TEMA 04:** Extensiones de los Modelos Gráficos.

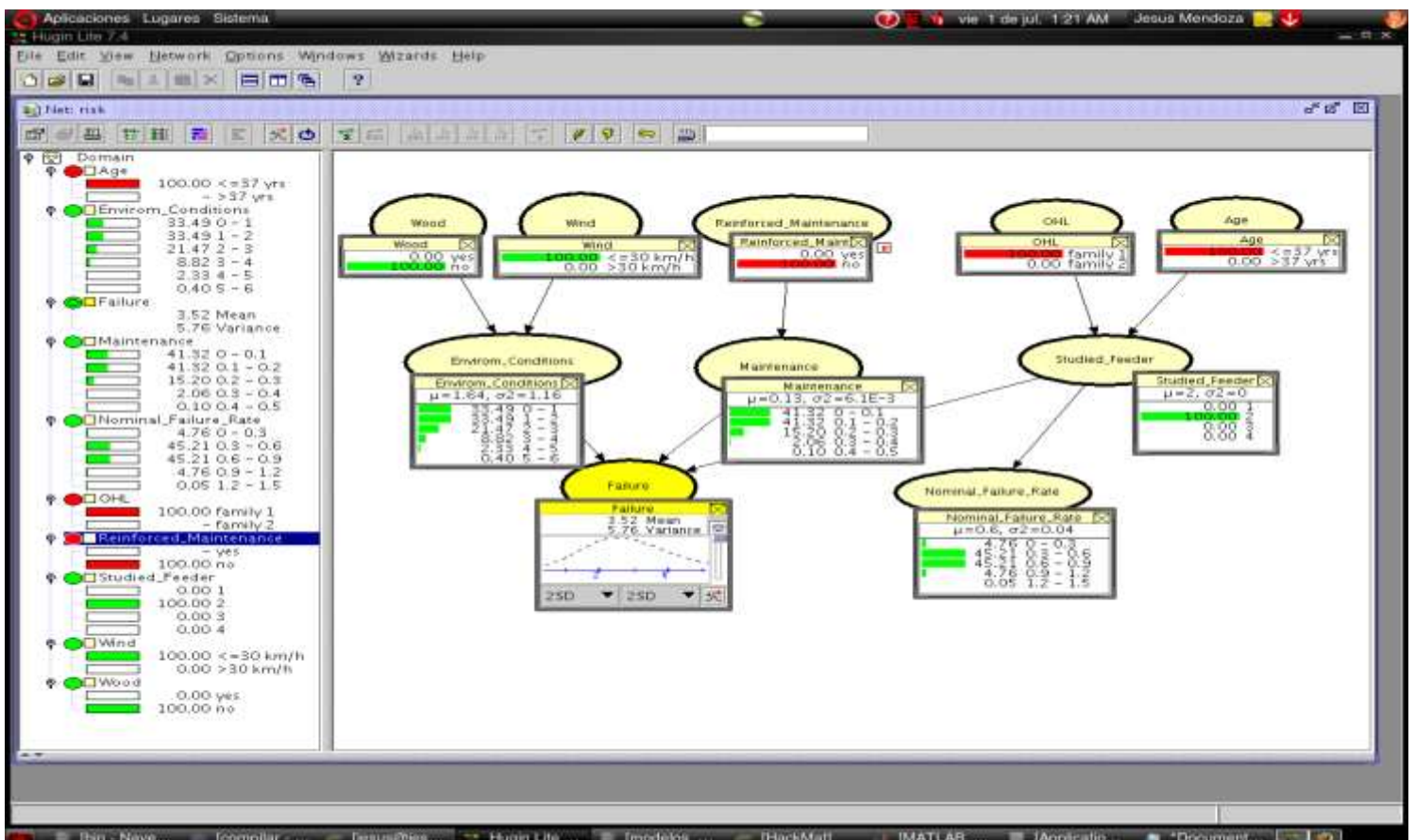
# Construcción de Modelos Probabilísticos

## TEMA 1



**Competencia:**

**Interpretar la construcción de modelos probabilísticos.**





# Desarrollo de los Temas



## Tema 01: Construcción de Modelos Probabilísticos

### MÉTODOS DE BÚSQUEDA



Muchos algoritmos para grafos necesitan un mecanismo de búsqueda para explorar los nodos y aristas de un grafo. Por ejemplo, entre otras cosas, los algoritmos de búsqueda pueden ser utilizados para obtener un camino entre dos nodos, o para buscar un bucle o ciclo en un grafo.

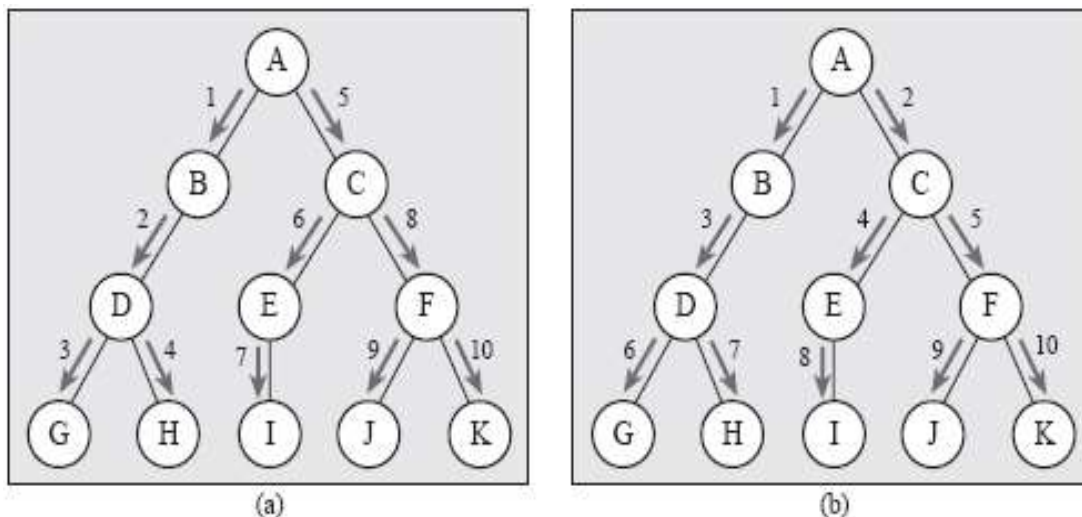
Estos métodos son la base para la construcción de los algoritmos introducidos en esta sección. La exploración de un grafo comienza en un nodo inicial y consiste en la definición de un criterio para moverse hacia adelante y hacia atrás a través de las aristas del grafo, pasando de un nodo a un nodo vecino en cada etapa.

Por tanto, la diferencia entre los distintos métodos de búsqueda radica en el criterio elegido para moverse de un nodo a otro.



- ❖ **Método de búsqueda en profundidad:** En cada etapa del método de búsqueda en profundidad se visita alguno de los vecinos no visitados del nodo actual donde los números indican el orden en que se visitan los nodos. En caso de que el nodo actual no tenga ningún vecino no visitado, el algoritmo vuelve atrás al nodo visitado anteriormente y el proceso de búsqueda continua hasta que todos los nodos han sido visitados.

- ❖ **Método de búsqueda en anchura:** El método de búsqueda en anchura visita los nodos del grafo capa a capa, comenzando en un nodo inicial y visitando, en la primera etapa todos los vecinos del nodo inicial. Después, se selecciona alguno de estos vecinos como nuevo nodo y se repite el proceso (ver Figura 4.46 (b), donde los números indican el orden en que se visitan los nodos).



**FIGURA 4.46.** Ilustración del método de búsqueda en profundidad (a) y de búsqueda en anchura (b). Los números indican el orden en que se visitan los nodos.

### Algoritmos de búsqueda de caminos

Dado un grafo  $G = (X, L)$ , se trata de encontrar un camino del nodo  $X_i$  al nodo  $X_j$ , en caso de que exista. En esta sección se introducen dos algoritmos de búsqueda de caminos basados en las dos estrategias anteriores. Para este propósito es más conveniente y eficiente utilizar la representación de un grafo por medio de los conjuntos de adyacencia. El grafo no dirigido de la Figura 4.47(a) puede ser representado por  $(X, L)$ , donde  $X$  es el conjunto de nodos  $\{A, B, C, D, E, F, G\}$  y  $L$  es el conjunto de aristas  $\{L_1, \dots, L_8\}$ . Sin embargo, desde un punto de vista computacional, la representación del grafo por medio de sus conjuntos de adyacencia es más adecuada:



$Ady(A) = \{B, C, D\},$

$Ady(B) = \{A, E\},$

$Ady(C) = \{A, F\},$

$Ady(D) = \{A, F\},$

$Ady(E) = \{B, G\},$

$Ady(F) = \{C, D, G\},$

$Ady(G) = \{E, F\}.$

Esta representación es más eficiente para los métodos de búsqueda pues evita tener que comprobar todas las aristas del grafo para elegir el siguiente nodo del proceso.

El grafo dirigido de la Figura 4.47(b) tiene los conjuntos siguientes de adyacencia:

$\text{Ady}(A) = \{B, C, D\}$ ,  $\text{Ady}(B) = \{E\}$ ,  $\text{Ady}(C) = \{F\}$ ,

$\text{Ady}(D) = \{F\}$ ,  $\text{Ady}(E) = \{G\}$ ,  $\text{Ady}(F) = \{G\}$ ,  $\text{Ady}(G) = \varnothing$ .

Otra propiedad importante de los conjuntos de adyacencia es que proporcionan una representación independiente del carácter dirigido o no dirigido del grafo. Por ejemplo, si nos diesen el grafo dirigido de la Figura 4.47 (b) y se quisiese realizar alguna operación de carácter no dirigido (obtener bucles, caminos no dirigidos, etc.), bastaría con considerar los conjuntos de adyacencia correspondientes al grafo no dirigido asociado.

Basándose en las dos técnicas de búsqueda descritas anteriormente, es posible definir de forma sencilla los siguientes algoritmos de búsqueda de caminos: búsqueda de caminos en profundidad y búsqueda de caminos en anchura.

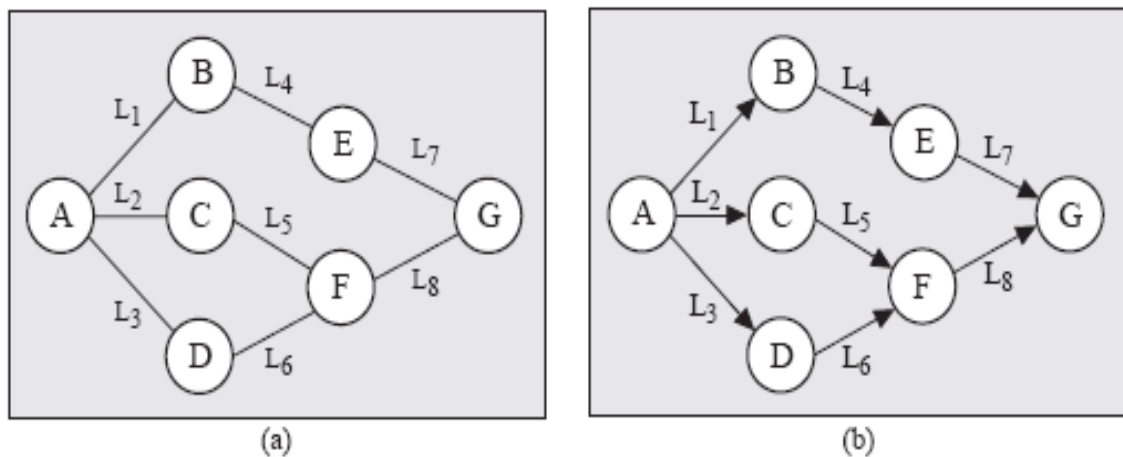


FIGURA 4.47. Ejemplo de un grafo no dirigido (a) y dirigido (b).

## Comprobando la Conexión de un Grafo

Los métodos de búsqueda descritos anteriormente también pueden utilizarse para comprobar si un grafo es conexo. La idea es realizar una búsqueda exhaustiva de los nodos del grafo, obteniendo el conjunto  $S$  de nodos que son alcanzables desde un nodo inicial. Si el grafo es conexo, entonces el del conjunto  $S$  contendrá todos los nodos del grafo, en caso contrario el del subconjunto de nodos  $S$  solo contendrá la componente conexas del grafo que contiene al nodo inicial.



Los Algoritmos pueden ser utilizados para realizar una búsqueda exhaustiva considerando el mismo nodo inicial y final, es decir, el conjunto Visitados resultante de la ejecución de estos algoritmos contendrá la componente conexas correspondiente a  $X_i$ .

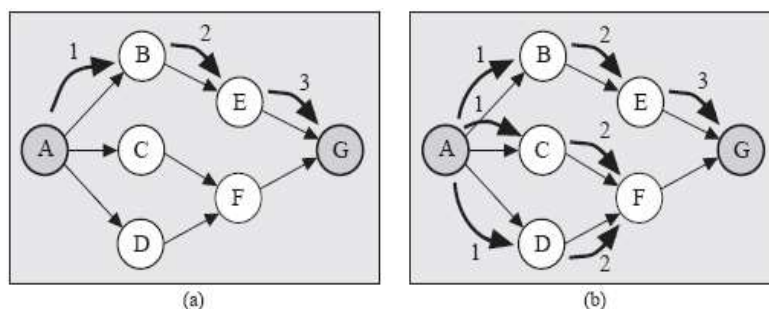


FIGURA 4.52. Búsqueda de un camino entre los nodos A y G con el algoritmo de búsqueda en profundidad (a) y en anchura (b).

## Búsqueda de componentes conexas

- Datos: Un grafo  $(X, \text{Ady})$ .
- Resultado: El conjunto de componentes conexas  $C$  de  $(X, \text{Ady})$ .
  1. Iniciación: Definir  $V_{\text{visitados}} = \emptyset$ ,  $C = \emptyset$ .
  2. Si  $X \setminus V_{\text{visitados}} = \emptyset$ , finalizar y devolver  $C$ ; en caso contrario, elegir un nodo de  $X_i \in X \setminus V_{\text{visitados}}$  e ir a la Etapa 3.
  3. Utilizar el Algoritmo para realizar una búsqueda exhaustiva del grafo  $(X, \text{Ady})$  comenzando en el nodo  $X_i$  y obtener el conjunto  $S$  de nodos visitados.
  4. Añadir  $S$  a  $C$ . Añadir a  $V_{\text{visitados}}$  todos los nodos en  $S$ . Ir a la Etapa 2.

Si el conjunto  $C$  contiene una sola componente conexa, entonces el grafo es conexo; en caso contrario, el grafo es inconexo y  $C$  contiene todas las componentes conexas del grafo.

## Búsqueda de bucles y ciclos



Los algoritmos de búsqueda de caminos pueden modificarse fácilmente para hallar bucles o ciclos en un grafo. En esta sección se muestran las modificaciones necesarias para adaptar el algoritmo de búsqueda en profundidad para esta tarea. Dado que el objetivo de este algoritmo es hallar un camino cerrado (un bucle o un ciclo), se puede utilizar el Algoritmo comprobando en cada etapa si hay algún nodo contenido en el camino que también esté contenido en la lista de nodos adyacentes del nodo actual. Los caminos cerrados resultantes serán bucles (si el grafo es no dirigido) o ciclos (si el grafo es dirigido). El algoritmo selecciona un nodo.

Construcción del modelo probabilístico. Una vez que se conoce un conjunto de variables relevantes para el problema a analizar, y que se ha adquirido suficiente información para su definición, el siguiente paso consiste en la definición de una función de probabilidad conjunta que describa las relaciones entre las variables. Éste es, quizás, el paso más crítico y difícil en el desarrollo de un sistema experto:

- a) Es crítico porque la bondad de los resultados del sistema experto depender de la precisión con que se haya definido la función de probabilidad conjunta, es decir, la calidad de los resultados no podrá superar a la calidad del modelo.

Por tanto, una incorrecta definición del modelo probabilístico redundará en un sistema experto que dará conclusiones erróneas y/o contradictorias.

- b) La estructura de la función de probabilidad conjunta (es decir, la estructura de dependencia e independencia entre las variables) no suele ser conocida en la práctica. Por tanto, habrá de ser inferida del conjunto de datos obtenidos previamente. Por tanto, la calidad del modelo tampoco podrá superar la calidad de los datos relevantes disponibles.

- c) La estructura del modelo probabilístico puede depender de un número muy elevado de parámetros que complican su definición. Cuanto mayor sea el número de parámetros más complicada será la asignación de valores numéricos concretos en el proceso de definición del modelo. En cualquier caso, esta asignación habrá de ser realizada por un experto, o estimada a partir de la información disponible.

## Criterios de separación gráfica

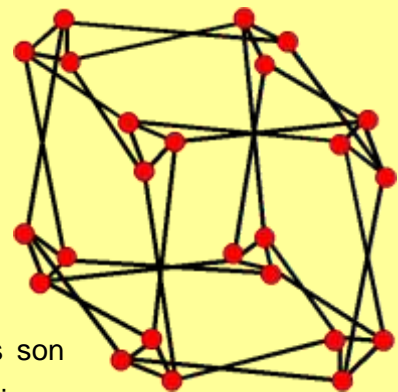


Los grafos son herramientas muy potentes para describir de forma intuitiva las relaciones de dependencia e independencia existentes en un conjunto de variables  $\{X_1, \dots, X_n\}$ . Por tanto, una forma de definir un modelo probabilístico es partir de un grafo que describa las relaciones existentes entre las variables.

Para comprobar cuáles, de entre todas las posibles relaciones de independencia condicional, son satisfechas por el grafo. Los criterios de separación gráfica son las reglas para entender cómo pueden codificarse dependencias e independencias en un grafo. Estos criterios dependen del tipo de grafo (dirigido o no dirigido) que se esté considerando.

## Separación en grafos no dirigidos

En muchas situaciones prácticas, las relaciones existentes entre un conjunto de variables  $\{X_1, \dots, X_n\}$  pueden ser representadas por un grafo no dirigido  $G$ . Cada variable puede ser representada por un nodo del grafo. Si dos variables son dependientes, esta relación puede representarse por un camino que conecte estos nodos. Por otra parte, si dos variables son independientes, entonces no deberá existir ningún camino que una estos nodos. De esta forma, el concepto de dependencia entre variables puede relacionarse con el concepto de conexión entre nodos.







De forma similar, si la dependencia entre las variables  $X$  e  $Y$  es indirecta, a través de una tercera variable  $Z$  (es decir, si  $X$  e  $Y$  son condicionalmente dependientes dada  $Z$ ), el nodo  $Z$  se representará de forma que no interseque todos los caminos entre  $X$  y  $Y$ , es decir,  $Z$  no es un conjunto de corte (en inglés, cutset) de  $X$  e  $Y$ .



Esta correspondencia entre dependencia condicional y separación en grafos no dirigidos constituye la base de la teoría de los campos de Markov (Isham (1981), Lauritzen (1982), Wermuth y Lauritzen (1983)), y ha sido caracterizada axiomáticamente de formas diversas (Pearl y Paz (1987)).

Para representar relaciones de independencia condicional por medio de grafos no dirigidos se necesita definir de forma precisa un criterio de separación apropiado, basándose en las ideas anteriormente expuestas. Este criterio se conoce como criterio de U-separación. A continuación se da una definición de este criterio y un algoritmo que permite su aplicación.

U-separación: Sean  $X$ ,  $Y$  y  $Z$  tres conjunto disjuntos de nodos de un grafo no dirigido  $G$ . Se dice que  $Z$  separa  $X$  e  $Y$  si y sólo si cada camino entre nodos de  $X$  y nodos de  $Y$  contiene algún nodo de  $Z$ . Cuando  $Z$  separe  $X$  e  $Y$  en  $G$ , y se denotar  $I(X, Y | Z) G$  para indicar que esta relación de independencia se deriva de un grafo  $G$ ; en caso contrario, se denotará por  $D(X, Y | Z) G$ , para indicar que  $X$  e  $Y$  son condicionalmente dependientes dada  $Z$ , en el grafo  $G$ .

Se dice que  $X$  es gráficamente independiente de  $Y$  dada  $Z$ . Si  $Z$  separa  $X$  e  $Y$ . Por tanto, el criterio de U-separación permite obtener la lista de relaciones de independencia asociadas a un grafo no dirigido.



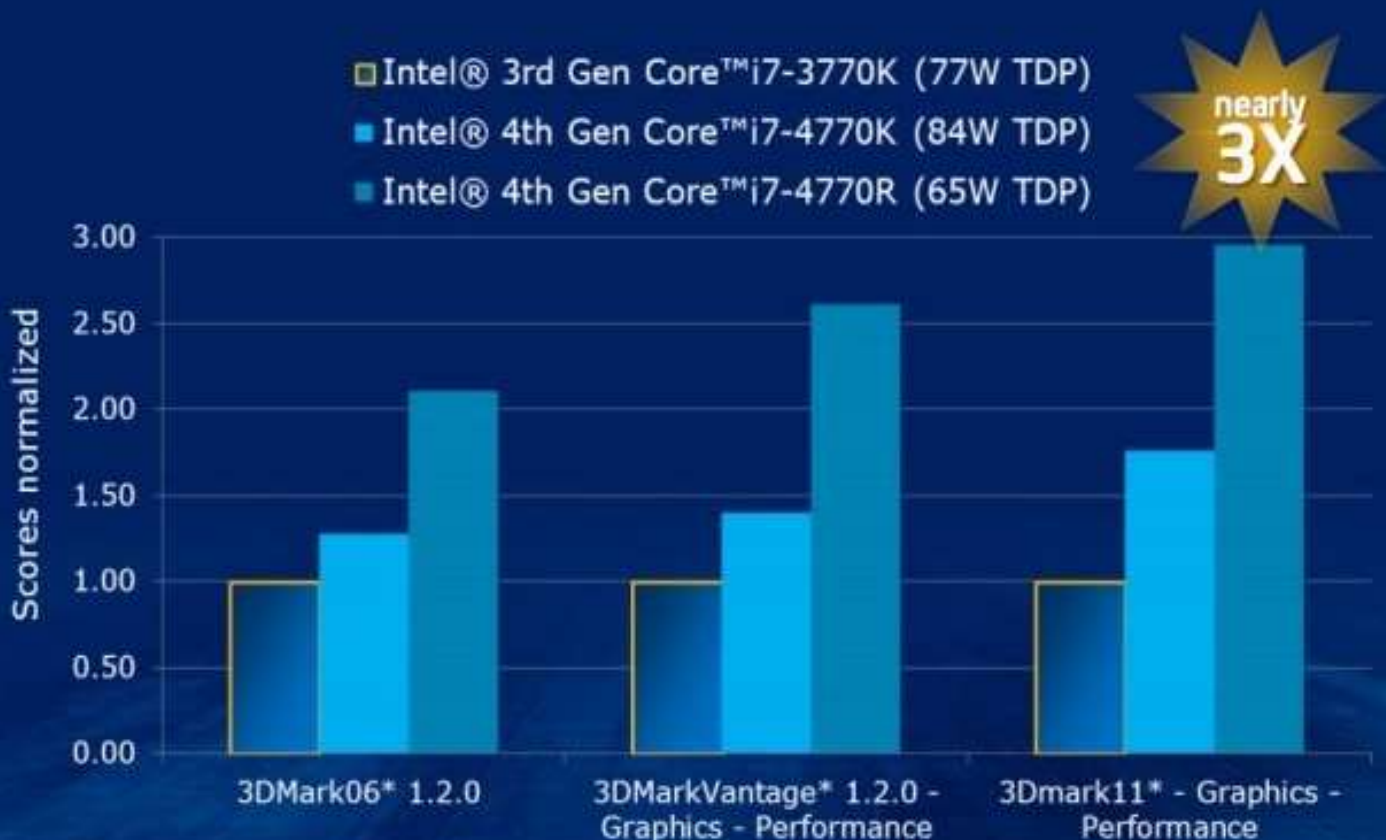
# *Modelos Definidos Gráficamente I*

## TEMA 2



### Competencia:

**Describir los modelos definidos gráficamente empleando sus técnicas.**





## Tema 02: Modelos Definidos Gráficamente I

### ALGUNAS PROPIEDADES DE LA INDEPENDENCIA CONDICIONAL

Hasta ahora se han introducido tres modelos distintos para definir relaciones de independencia condicional: modelos probabilísticos, modelos gráficos no dirigidos, y modelos gráficos dirigidos. En esta sección se analizan algunas propiedades de la independencia condicional que cumplen algunos de estos modelos.



Estas propiedades permiten obtener nuevas relaciones de independencia a partir de un conjunto inicial de relaciones de independencia, dado por uno de estos modelos. Por ejemplo, dada la función de probabilidad conjunta  $p(x_1, \dots, x_n)$  de un conjunto de variables  $\{X_1, \dots, X_n\}$ .

### Modelos de dependencia

**Grafoide:** Un grafoide es un conjunto de relaciones de independencia que es cerrado con respecto a las propiedades de simetría, descomposición, unión débil, contracción e intersección.

**Semigrafoide.** Un semigrafoide es un conjunto de relaciones de independencia que es cerrado con respecto a las propiedades de simetría, descomposición, unión débil y contracción.

Por tanto, un grafoide debe satisfacer las cinco primeras propiedades, mientras que un semigrafoide debe satisfacer sólo las cuatro primeras.

Dada una lista inicial de independencias, un grafo, o una función de probabilidad conjunta, siempre es posible determinar que relaciones de independencia se cumplen en el modelo y, por tanto, determinar su estructura cualitativa. Por tanto, estos tipos de modelos definen clases particulares de los denominados modelos de dependencia.

**Modelo de Dependencia.** Cualquier modelo  $M$  de un conjunto de variables  $\{X_1, \dots, X_n\}$  mediante el cual se pueda determinar si la relación  $I(X, Y | Z)$  es o no cierta, para todas las posibles ternas de subconjuntos  $X, Y$  y  $Z$ , se denomina modelo de dependencia.

**Modelo de dependencia probabilístico:** Un modelo de dependencia  $M$  se denomina probabilístico si contiene todas las relaciones de independencia dadas por una función de probabilidad conjunta  $p(x_1, \dots, x_n)$ .

**Modelo de dependencia probabilístico no extremo:** Un modelo de dependencia probabilístico no extremo es un modelo de dependencia probabilístico obtenido de una función de probabilidad no extrema, o positiva; es decir,  $p(x_1, \dots, x_n)$  toma valores en el intervalo abierto  $(0, 1)$ .

Dado que todas las funciones de probabilidad satisfacen las cuatro primeras propiedades de independencia condicional, todos los modelos de dependencia probabilísticos son semigrafoides.



Por otra parte, dado que solo las funciones de probabilidad no extremas satisfacen la propiedad de intersección, solo los modelos de dependencia probabilísticos no extremos son grafoides.



**Modelo de dependencia compatible con una probabilidad:** Un modelo de dependencia  $M$  se dice compatible con una función de probabilidad  $p(x_1, \dots, x_n)$  si todas las relaciones de independencia derivadas  $M$  son también satisfechas por  $p(x_1, \dots, x_n)$ .

Obsérvese que un modelo de dependencia compatible con una probabilidad es aquel que puede obtenerse de una función de probabilidad conjunta  $p(x_1, \dots, x_n)$ , pero sin necesidad de ser completo, es decir, no tienen por que contener todas las relaciones de independencia que pueden obtenerse de  $p(x_1, \dots, x_n)$ .

Dado que toda función de probabilidad cumple las cuatro primeras propiedades de la independencia condicional, si un modelo de dependencia  $M$  es compatible con una función de probabilidad  $p(x_1, \dots, x_n)$ , entonces el menor semigrafoide generado por  $M$  también debe ser compatible con  $p(x_1, \dots, x_n)$ . Por tanto, un problema interesante desde el punto de vista práctico es calcular el menor semigrafoide generado por un modelo de dependencia  $M$ .



## Construcción de un modelo probabilístico



El problema de construir una función de probabilidad para un conjunto de variables puede simplificarse notablemente considerando una factorización de la probabilidad como producto de funciones de probabilidad condicionada más sencillas. El grado de implicación dependerá de la estructura de independencia (incondicional o condicional) existente entre las variables del modelo. Por tanto, para encontrar una factorización apropiada del modelo probabilístico, primero se necesita conocer su estructura de independencia.

**Esta estructura de independencia (modelo de dependencia) caracteriza la estructura cualitativa de las relaciones entre las variables.**

Por ejemplo, se necesita definir que variables son independientes y/o condicionalmente independientes de otras y cuáles no. La estructura de independencia y, por tanto, la factorización asociada al modelo probabilístico, puede ser obtenida de varias formas:

1. **Modelos definidos gráficamente:** Como se ha visto en las secciones anteriores, las relaciones existentes entre las variables de un conjunto pueden ser descritas mediante un grafo. Posteriormente, utilizando un criterio de separación apropiado, se puede obtener el conjunto de relaciones de independencia asociado. Estos modelos de dependencia se conocen como modelos definidos gráficamente, y tienen como ejemplos más importantes a las redes de Markov, y las redes Bayesianas.

*Las tareas de comprobar la validez de un grafo, entender sus implicaciones, y modificarlo de forma apropiada han de ser realizadas partiendo de la comprensión de las relaciones de dependencia e independencia existentes en el conjunto de variables.*



2. **Modelos definidos por listas de independencias:** Los grafos son herramientas muy útiles para definir la estructura de independencia de un modelo probabilístico. Una descripción alternativa a los modelos graficos consiste en utilizar directamente un conjunto  $M$  de relaciones de independencia que describan las relaciones entre las variables. Este conjunto puede ser definido por un experto a partir de sus opiniones sobre las relaciones entre las variables del modelo. Cada una de las independencias del conjunto indica que variables contienen información relevante sobre otras y cuando el conocimiento de algunas variables hace que otras sean irrelevantes para un conjunto de variables dado.

*Este conjunto inicial de independencias puede ser completado incluyendo aquellas otras que cumplan una serie de propiedades de independencia condicional. El conjunto resultante puede ser finalmente utilizado para obtener una factorización de la función de probabilidad del modelo.*



Los modelos resultantes se conocen como modelos definidos por listas de relaciones de independencia.

**3. Modelos definidos condicionalmente:** Como alternativa a los modelos gráficos y los modelos dados por listas de relaciones de independencia, la estructura cualitativa de un modelo probabilístico puede venir dada por un conjunto de funciones de probabilidad marginales y condicionadas.

$$P = \{p_1(u_1|v_1), \dots, p_m(u_m|v_m)\}$$

Sin embargo, las funciones de este conjunto no pueden definirse libremente, sino que han de satisfacer ciertas relaciones para ser compatibles y definir un único modelo probabilístico. Una ventaja de utilizar modelos gráficos, o modelos definidos por listas de independencias, para construir un modelo probabilístico es que estos modelos definen una factorización de la función de probabilidad como producto de funciones de probabilidad condicionada que determinan la estructura cualitativa del modelo probabilístico. Normalmente, estas funciones condicionadas contienen un número menor de variables que la función de probabilidad conjunta y, por tanto, el proceso de definición del modelo probabilístico es más sencillo.



Una vez que se conoce la estructura cualitativa del modelo probabilístico (la factorización de la función de probabilidad), la estructura cuantitativa de un modelo particular se define mediante la asignación de valores numéricos a los parámetros asociados a las funciones de probabilidad condicionada que intervienen en la factorización del modelo. Estos valores han de ser definidos por algún experto, o estimados a partir de un conjunto de datos. Por tanto, si la estructura cualitativa del modelo es desconocida, que es el caso habitual en la práctica, entonces tanto la estructura cualitativa, como la cuantitativa (los parámetros) han de ser estimadas a partir del conjunto de datos disponible (una base de datos, etc.).



**Como resumen de todo lo anterior, la construcción de un modelo probabilístico puede ser realizada en dos etapas:**

1. Factorizar la función de probabilidad mediante un producto de funciones de probabilidad condicionada. Esta factorización puede obtenerse de tres formas distintas:
2. Estimar los parámetros de cada una de las funciones de probabilidad condicionada resultantes.



Este proceso se ilustra de modo esquemático en la Figura 5.9. En este diagrama, una línea continua de un rectángulo A a un rectángulo B significa que cada miembro de A es también un miembro de B, mientras que una línea discontinua significa que algunos, pero no necesariamente todos, los miembros de A son miembros de B. El camino más simple para definir un modelo probabilístico es comenzar con un grafo que se supone describe la estructura de dependencia e independencia de las variables.



A continuación, el grafo puede utilizarse para construir una factorización de la función de probabilidad de las variables. De forma alternativa, también puede comenzarse con una lista de relaciones de independencia y, a partir de ella, obtener una factorización de la función de probabilidad. La factorización obtenida determina los parámetros necesarios para definir el modelo probabilístico. Una vez que estos parámetros han sido definidos, o estimados a partir de un conjunto de datos, la función de probabilidad que define el modelo probabilístico vendrá como el producto de las funciones de probabilidad condicionada resultantes.

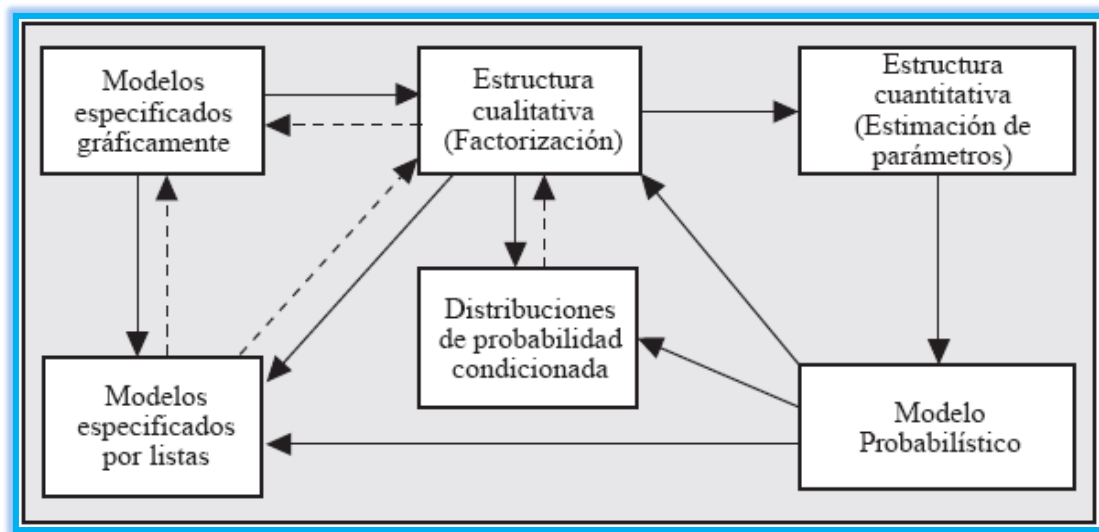


FIGURA 5.9. Diagrama mostrando las formas alternativas de definir un modelo probabilístico.

Por otra parte, si se conoce la función de probabilidad que define un modelo probabilístico (que no es el caso habitual en la práctica), se puede seguir el camino inverso y obtener varias factorizaciones distintas. También se puede obtener la lista de independencias correspondiente al modelo comprobando cuales de todas las posibles relaciones de independencia de las variables son verificadas por la función de probabilidad. A partir del conjunto de independencias obtenido, también puede construirse una factorización de la familia paramétrica que contiene a la función de probabilidad dada.



# Modelos Definidos Gráficamente II

## TEMA 3



### Competencia:

**Aplicar las técnicas y herramientas formales de análisis de sistemas desde los modelos definidos gráficamente.**





## Tema 03: Modelos Definidos Gráficamente II



Se ha visto que el funcionamiento de un sistema experto probabilístico depende de la correcta definición del correspondiente modelo, que está caracterizado por la función de probabilidad conjunta de las variables. También se ha visto que la estructura general de una función de probabilidad conjunta involucra un excesivo número de parámetros. Por esta razón, se presentaron algunos modelos probabilísticos simplificados, que eran obtenidos imponiendo ciertas hipótesis de independencia globales sobre las variables. Sin embargo, estos modelos son restrictivos y solamente aplicables a problemas del tipo “enfermedades-síntomas”.

En este capítulo se desarrolla la forma de obtener modelos probabilísticos más generales por medio de grafos. La idea básica consiste en utilizar grafos (no dirigidos o dirigidos) para construir un modelo de dependencia que represente la estructura cualitativa del modelo probabilístico. De esta forma, los modelos resultantes son generales, pues se crean a partir de un modelo de dependencia “arbitrario”, y no de uno impuesto inicialmente. Conjuntos de variables son incondicionalmente o condicionalmente dependientes o independientes.



Cada modelo probabilístico tiene asociado un modelo de dependencia  $M$ , que puede ser obtenido generando todas las relaciones de independencia condicional posibles para un conjunto de variables dado, y comprobando cuales de ellas se satisfacen para la función de probabilidad. Por ejemplo, si  $X$ ,  $Y$  y  $Z$  son tres subconjuntos disjuntos y  $p(x|y, z) = p(x|z)$ , para cada combinación de valores de  $x$ ,  $y$  y  $z$ , entonces se verifica la relación de independencia  $I(X, Y | Z)$  y se puede concluir que  $X$  e  $Y$  son condicionalmente independientes dado  $Z$ . Por otra parte, si  $p(x|y, z) = p(x|z)$  para algunos valores  $x$ ,  $y$ ,  $z$ , entonces  $X$  e  $Y$  son condicionalmente dependientes dado  $Z$ .



Por tanto, una función de probabilidad contiene una descripción completa (cuantitativa y cualitativa) de las relaciones entre las variables, mientras que el modelo de dependencia  $M$  asociado solo contiene una descripción cualitativa. Por tanto, el término modelo de dependencia probabilístico se refiere únicamente a un modelo de dependencia asociado a una función de probabilidad.

Por otra parte, un modelo de dependencia puede ser definido de forma alternativa mediante un grafo (dirigido o no dirigido), una lista de relaciones de independencia, o un conjunto de funciones de probabilidad condicionada. Estas tres alternativas determinan tres metodologías diferentes para construir un modelo de dependencia:

- Modelos definidos graficamente.
- Modelos definidos por listas de independencias.
- Modelos definidos condicionalmente.



Estas tres metodologías son más generales que los modelos presentados y pueden ser aplicadas, no solo a problemas de diagnóstico médico (problemas tipo “síntoma-enfermedad”), sino también a problemas más generales. Estas metodologías requieren ciertos conceptos previos, se ha visto que un conjunto de variables  $X_1, \dots, X_n$  y sus relaciones pueden ser representados mediante un grafo, asociando cada variable a un nodo y cada relación entre variables a una arista entre los nodos correspondientes. Por tanto, los términos nodo y variable se utilizan de forma sinonimia.



En algunas ocasiones, el orden de las variables (es decir, la dirección de las aristas) es importante en el grafo (grafos dirigidos) y en otras no (grafo no dirigido). Las representaciones gráficas tienen la ventaja de mostrar explícitamente las relaciones entre las variables y conservar estas relaciones de forma cualitativa (es decir, para cualquier valor numérico de los parámetros).

Los modelos gráficos son también más intuitivos y fáciles de entender. Se analizaron dos criterios gráficos de separación distintos para obtener las relaciones de independencia definidas por los grafos dirigidos y los no dirigidos. Según esta distinción, los modelos definidos bracammente pueden ser clasificados en dos grupos, dependiendo del tipo de grafo que se utilice:

- ✓ Modelos de dependencia definidos por grafos no dirigidos.
- ✓ Modelos de dependencia definidos por grafos dirigidos.



Aunque existe un tercer tipo de modelos gráficos que pueden ser representados por grafos mixtos (grafos que contienen aristas dirigidas y no dirigidas).

Se ha utilizado el termino dependencia en las definiciones anteriores para enfatizar que un grafo sólo puede definir la estructura cualitativa del modelo. Una vez que se conoce esta estructura cualitativa, puede construirse una factorización de la función de probabilidad e identificarse el conjunto de parámetros que definen el modelo. Los valores numéricos de los parámetros pueden ser dados por un experto, o estimados a partir de un conjunto de datos disponibles.

## Algunas definiciones y problemas

Mapa perfecto. Un grafo  $G$  se dice que es un mapa perfecto de un modelo de dependencia  $M$  si cada relación de independencia obtenida de  $G$  también puede ser obtenida de  $M$  y viceversa, es decir,

$$I(X, Y|Z)_M \Leftrightarrow I(X, Y|Z)_G \Leftrightarrow Z \text{ separa } X \text{ de } Y.$$

Dependiendo del carácter dirigido o no dirigido del grafo  $G$ , los mapas perfectos se denominan mapas perfectos dirigidos o no dirigidos, respectivamente.

Ocho posibles grafos no dirigidos con tres variables.

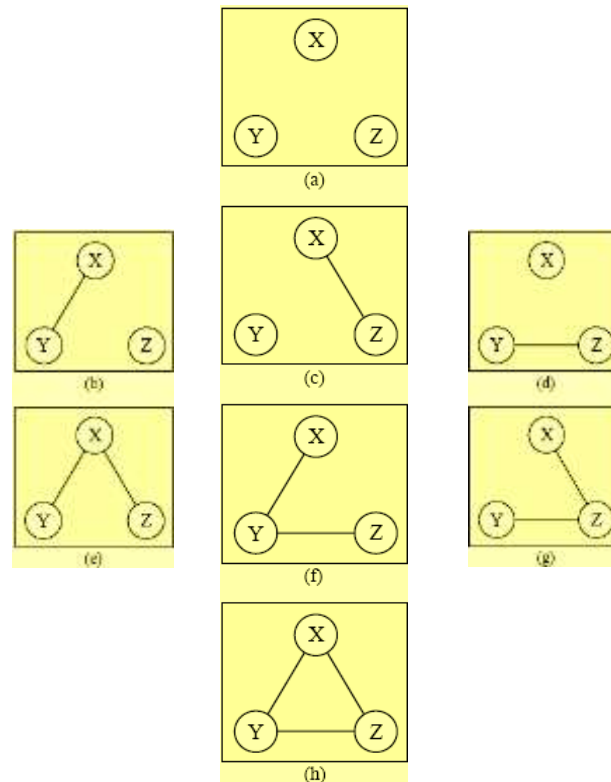


FIGURA 6.1. Ocho posibles grafos no dirigidos con tres variables.



El modelo de dependencia  $M$  del Ejemplo 6.1 tiene un mapa perfecto dirigido, a pesar de que no posee ningún mapa perfecto no dirigido. Se deja como ejercicio para el lector demostrar que el grafo dirigido mostrado en la Figura 6.2 es un mapa perfecto dirigido de  $M$ . En este caso, los grafos dirigidos son más potentes que los no dirigidos. Sin embargo, no todo modelo de dependencia posee un mapa perfecto dirigido. El ejemplo siguiente muestra uno de estos modelos.

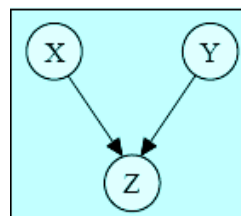


FIGURA 6.2. Mapa perfecto dirigido del modelo de dependencia  $M$  en (6.1)

**No existe ningún grafo dirigido acíclico  $D$  que sea mapa perfecto del modelo de dependencia  $M$ .**

En los casos en los que no existe un mapa perfecto, es necesario asegurarse de que el modelo gráfico que se utilice no posea ninguna independencia que no esté contenida en el modelo, y que el número de independencias del modelo que no sean reproducidas por el grafo sea mínimo. Esto motiva las siguientes definiciones.

**Mapa de independencia.** Un grafo  $G$  se dice que es un mapa de independencia (I - mapa) de un modelo de dependencia  $M$  si

$$I(X, Y|Z)_G \Rightarrow I(X, Y|Z)_M$$

Es decir, si todas las relaciones de dependencia derivadas de  $G$  son verificadas por  $M$ . Obsérvese que un I-mapa  $G$  de un modelo de dependencia  $M$  incluye algunas de las independencias de  $M$ , pero no necesariamente todas. Entonces, se tiene

$$I(X, Y|Z)_G \Rightarrow I(X, Y|Z)_M$$



Por tanto, cada modelo de dependencia tiene asociados un I-mapa y un D-mapa triviales. Por ejemplo, cualquier grafo totalmente inconexo es un D-mapa trivial y cualquier grafo completo es un I-mapa trivial de cualquier modelo de dependencia. De esta forma, para que un grafo sea un mapa perfecto de un modelo, ha de ser simultáneamente un

I-mapa y un D-mapa de ese modelo.

**I-mapa minimal:** Se dice que un grafo  $G$  es un I mapa minimal de un modelo de dependencia  $M$  si es un I-mapa de  $M$ , pero pierde esta propiedad cuando se elimina una cualquiera de sus aristas.

A pesar de que los modelos de dependencia y las representaciones gráficas tienen numerosas aplicaciones mas allá de la probabilidad, el interés principal de este libro es la construcción de modelos probabilísticos y, por tanto, estamos interesados en conocer la relación existente entre las representaciones gráficas y las funciones de probabilidad, es decir, la relación existente entre las nociones formales de dependencia probabilística y la estructura topológica de un de un grafo.

Una razón importante para representar la estructura de dependencia de un modelo mediante un grafo es que comprobar la conexión de un conjunto de variables en un grafo, es más fácil que comprobar la independencia condicional de un conjunto de variables utilizando las formulas de la Probabilidad.

Un D- mapa garantiza que todos los nodos que estén conectados en el grafo serán por tanto dependientes; sin embargo, el grafo puede ocasionalmente representar desconectados algunos conjuntos de variables dependientes.

### Modelos de dependencia gráficos no dirigidos



En esta sección se analiza la forma de definir modelos de dependencia utilizando grafos no dirigidos. Nuestro objetivo es encontrar un grafo que reproduzca tantas independencias asociadas a un modelo probabilístico como sea posible. Se comienza con el problema de representar estos modelos por medio de mapas perfectos los mapas y, a continuación, se introduce un clase importante de modelos probabilísticos definidos por grafos no dirigidos. Estos modelos se conocen por redes de Markov.

### De modelos a grafos no dirigidos

En esta sección se analiza el problema de representar modelos probabilísticos utilizando grafos no dirigidos, es decir, se desea encontrar el grafo correspondiente a un modelo de dependencia probabilístico. Como ya se ha visto, no todos lo modelos probabilísticos de dependencia pueden ser representados por mapas perfectos no dirigidos.



# *Extensiones de los Modelos Gráficos*

## TEMA 4



### Competencia:

**Emplear las extensiones de los modelos gráficos para desarrollar de manera técnica las probabilidades.**







## Tema 04: Extensiones de los Modelos Gráficos

### MODELOS DEFINIDOS POR MULTÍGRAFOS

#### Interpretación de independencias en un multígrafo

El primer problema relacionado con los modelos definidos por multígrafos es la interpretación gráfica de sus independencias. Las redes Bayesianas y de Markov son I-mapas de un cierto modelo de dependencia asociado al modelo probabilístico correspondiente. Entonces, todas las independencias condicionales contenidas en el grafo también son independencias del modelo correspondiente.



Por tanto, será cierta en un multígrafo una relación de independencia cualquiera si es cierta en alguno de los grafos que componen el multígrafo; en caso contrario será falsa. Por tanto, el criterio gráfico de separación para multígrafos consiste en la aplicación del criterio de U - separación en los grafos no dirigidos que compongan el multígrafo y el criterio de D-separación en los dirigidos.

#### Reducción del conjunto de grafos

El segundo problema de estos modelos es el de la redundancia en un multígrafo. En algunos casos, todas las independencias implicadas por un grafo del modelo pueden ser obtenidas a partir de los demás grafos. Por ejemplo, Shachter (1990b) introdujo algunas transformaciones gracias que permiten simplificar la estructura de los grafos eliminando independencias redundantes.



En algunos casos, el conjunto de grafos puede ser reducido a un conjunto menor que es una representación más simple y eficiente del modelo.

**Grafos redundantes.** Dados dos grafos  $G_1$  y  $G_2$ , se dice que  $G_1$  es redundante dado  $G_2$  si el conjunto de relaciones de independencia contenidas en  $G_1$  está contenido en  $G_2$ .

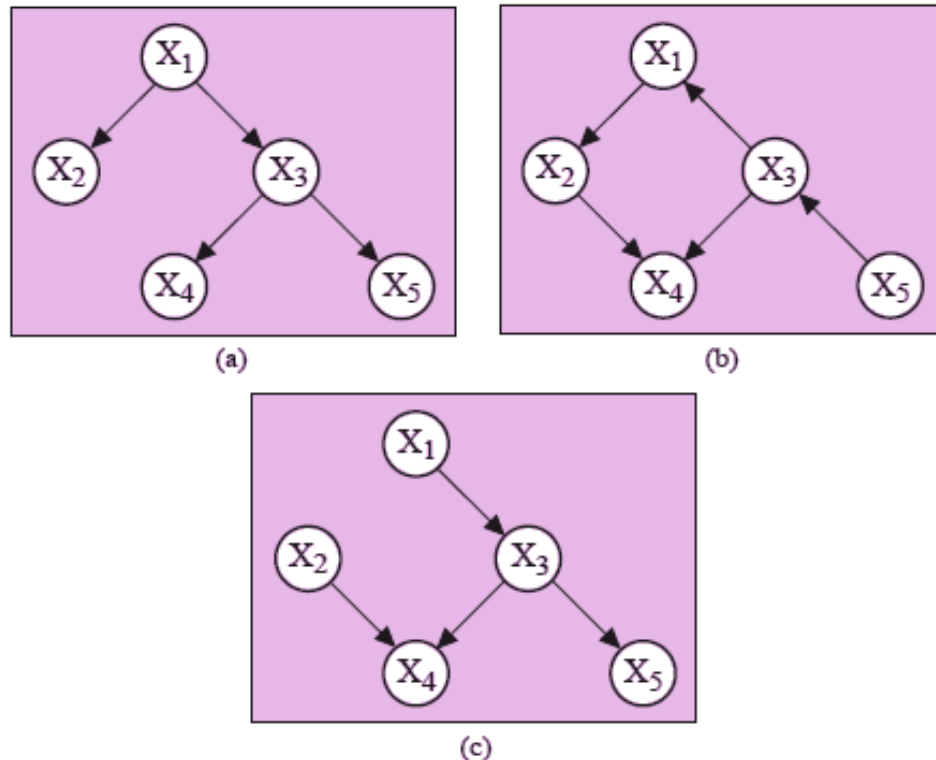


FIGURA 7.3. Tres grafos dirigidos acíclicos que definen un multigrafo.

**El teorema siguiente muestra las condiciones para que dos grafos dirigidos sean redundantes.**

**Redundancia en multígrafos dirigidos.** Sean  $D_1$  y  $D_2$  dos grafos dirigidos acíclicos sobre el mismo conjunto de variables  $X$ , y sean  $G_1$  y  $G_2$  los grafos no dirigidos asociados respectivos. Entonces,  $D_2$  es redundante dado  $D_1$  si (a)  $G_2$  está contenido en  $G_1$ , (b) cada v-estructura de  $D_1$  está también contenida en  $D_2$ , y (c) cada v-estructura  $(X_i, X_j, X_k)$  de  $D_2$  está también contenida en  $D_1$  siempre que  $G_1$  contenga el camino  $X_i - X_j - X_k$ .

## Modelos definidos por listas de independencias

Las listas de independencias constituyen una alternativa a los modelos gráficos para la construcción de modelos probabilísticos. Esta lista puede venir dada directamente por un experto en el tema a analizar, y representa las relaciones existentes entre las variables del modelo.

En esta sección se analiza la relación entre una relación de independencia en un modelo probabilístico y una factorización de la función de probabilidad correspondiente.

Esta relación puede resumirse del modo siguiente:

- ❖ Siempre se puede encontrar una factorización que contiene una relación de independencia dada.
- ❖ Una factorización puede implicar una o más relaciones de independencia.



De una relación de independencia a una factorización. Considérese el conjunto de variables  $\{X_1, X_2, X_3, X_4\}$  y supóngase que cumplen la relación de independencia  $I(X_1, X_2 | X_3)$ . La función de probabilidad correspondiente puede escribirse como

$$\begin{aligned} p(x_1, x_2, x_3, x_4) &= p(x_2, x_3)p(x_1|x_2, x_3)p(x_4|x_1, x_2, x_3) \\ &= p(x_2, x_3)p(x_1|x_3)p(x_4|x_1, x_2, x_3). \end{aligned} \quad \text{.....(7.14)}$$

Donde la primera igualdad se ha obtenido considerando la partición de las variables  $\{\{X_2, X_3\}, X_1, X_4\}$  y aplicando la regla de la cadena a la función de probabilidad  $p(x)$ , y la segunda igualdad se ha obtenido utilizando la relación de independencia  $I(X_1, X_2 | X_3)$ , que implica  $p(x_1 | x_2, x_3) = p(x_1 | x_3)$ .

**Por tanto, cualquier función de probabilidad que factorice según (7.14) contiene, al menos, la relación de independencia  $I(X_1, X_2 | X_3)$ .**

Obsérvese que la función de probabilidad podría contener también otras relaciones de independencia derivadas de los axiomas de la probabilidad (por ejemplo, la relación de independencia simétrica  $I(X_2, X_1 | X_3)$ ). Por tanto, la lista de independencias formada por una única relación de independencia es un I-mapa del modelo probabilístico resultante.

Existen listas de independencia que contienen varias relaciones de independencia y que pueden definir una única factorización de forma colectiva. Un ejemplo de ello lo constituyen las listas causales. Dado el conjunto de variables  $X = \{X_1, \dots, X_n\}$ , una lista causal definida sobre  $X$  es un conjunto de relaciones de independencia de la forma  $\{I(Y_1, B_1 \setminus S_1 | S_1), \dots, I(Y_n, B_n \setminus S_n | S_n)\}$ , donde  $(Y_1, \dots, Y_n)$  es una permutación de  $\{X_1, \dots, X_n\}$  y  $S_i \subset B_i = \{Y_1, \dots, Y_{i-1}\}$ .

**Esta lista define la siguiente:**

Factorización de la función de probabilidad

$$p(y_1, \dots, y_n) = \prod_{i=1}^n p(y_i | s_i), \quad \dots\dots\dots(7.15)$$

Que incluye todas las relaciones de independencia de la lista causal.

## Modelos probabilísticos multifactorizados

Hemos visto que la definición de una función de probabilidad mediante multigrafos y listas de relaciones de independencia se reduce a hallar la función de probabilidad compatible con un conjunto dado de factorizaciones. Por tanto, estos dos modelos son casos especiales de un tipo de modelos más generales conocido como modelos probabilísticos multifactorizados.



Modelos probabilísticos multifactorizados: Un modelo probabilístico multifactorizado sobre un conjunto de variables  $X = \{X_1, \dots, X_n\}$ , es un conjunto de factorizaciones compatibles obtenidas aplicando la regla de la cadena.

$$P = \{P^\ell, \ell = 1, \dots, m\}, \dots\dots\dots(7.28)$$

### Modelos multinomiales multifactorizados

#### Estructura paramétrica de una función de probabilidad

Considérese el conjunto de variables discretas  $\{X_1, \dots, X_n\}$ , donde la variable  $X_i$  puede tomar los valores  $\{0, \dots, r_i\}$ . Dado que las funciones de probabilidad condicionada  $p_\sigma(y_\sigma | s_\sigma)$ , que definen las factorizaciones de la función de probabilidad, pueden ser consideradas como familias paramétricas, una representación apropiada de los parámetros del modelo probabilístico asociado a la factorización Bésima viene dada por:

$$\theta_{ijs}^\ell = p(Y_i^\ell = j | S_i^\ell = s), \quad j \in \{0, \dots, r_i^\ell\}, \dots\dots\dots(7.30)$$

Donde  $s$  es una realización de  $S_\sigma$ . Por tanto, el primer subíndice de  $\theta_{ijs}$  se refiere al número del nodo, el segundo subíndice se refiere al estado del nodo y los subíndices restantes se refieren a la realización de  $S_\sigma$ . Dado que los parámetros están asociados a probabilidades, han de satisfacer las igualdades

$$\sum_{j=0}^{r_i^\ell} \theta_{ijs}^\ell = 1, \quad \ell = 1, \dots, m,$$

Para cada  $i$  y  $s$ . Por tanto, uno de los parámetros puede escribirse como uno menos la suma de los restantes. Por ejemplo,

$$\theta_{ir_i s}^\ell = 1 - \sum_{j=0}^{r_i^\ell - 1} \theta_{ijs}^\ell, \quad \ell = 1, \dots, m. \dots\dots\dots(7.31)$$

**El conjunto de parámetros  $\theta\sigma$  se denota por  $\Theta\sigma$ .**

La probabilidad de cualquier realización de las variables  $\{x_1, \dots, x_n\}$  es un monomio en los parámetros que definen el modelo probabilístico de grado menor o igual que el número de variables. Sin embargo es un polinomio de primer grado en cada uno de los parámetros.

Demostración: Se tiene que la probabilidad de una realización

$(x_1, \dots, x_n)$ , es

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | s_i) = \prod_{i=1}^n \theta_{ix_i s_i}.$$

Obsérvese que todos los parámetros que intervienen en el producto anterior están asociados a variables distintas. Por tanto  $p(x_1, \dots, x_n)$  es un monomio de grado menor o igual que el número de variables. Obsérvese también que  $p(x_1, \dots, x_n)$  puede resultar un polinomio si sólo se considera el conjunto de parámetros libres. Para ello solo se necesita reemplazar los parámetros  $\theta_{ir_i s_i}$  por

$$\theta_{ir_i s_i} = 1 - \prod_{j=0}^{r_i-1} \theta_{ijs_i}.$$



Esta substitución crea tantos monomios nuevos como cardinalidad tenga la variable  $X_i$ , pero cada uno de los monomios resultantes sigue siendo de primer grado en cada uno de los parámetros. El corolario siguiente determina la estructura algebraica de las probabilidades

marginales asociadas a un modelo probabilístico.

La probabilidad marginal de cualquier conjunto de nodos  $Y \subset X$  es un polinomio en los parámetros que definen el modelo probabilístico de grado menor o igual que el número de variables. Sin embargo, es un polinomio de primer grado en cada uno de los parámetros.

### El problema de la compatibilidad

El análisis de la estructura paramétrica de las probabilidades, introducido en la sección anterior, permite resolver el problema de la compatibilidad de los modelos multifactorizados, es decir, permite obtener la familia de funciones de probabilidad compatible con el conjunto de factorizaciones dado en (7.29). Obsérvese que siempre existe una solución trivial para este problema, ya que el modelo de independencia total cumple todas las relaciones de independencia posibles



Sin embargo, se está interesado en obtener una función de probabilidad que cumpla las relaciones de independencia necesarias, pero que incluya el mínimo número posible de independencias adicionales.

La idea del método propuesto por Castillo, Gutierrez y Hadi (1996b) es la de elegir una de las factorizaciones, por ejemplo  $P_1$ , y designarla como la factorización de referencia de la función de probabilidad. Los parámetros asociados,  $\Theta_1$ , también se denominan parámetros de referencia. Una vez que la factorización de referencia ha sido fijada, el problema de la compatibilidad puede ser resuelto calculando las restricciones sobre los parámetros de referencia para que la función de probabilidad pueda ser factorizada según el resto de factorizaciones.

### Modelos normales multifactorizados

El problema de compatibilidad asociado al conjunto de factorizaciones de un modelo normal multifactorizado se reduce al problema de encontrar la matriz de covarianzas de la variable aleatoria multidimensional que sea compatible con las factorizaciones dadas, o con las relaciones de independencia implicadas por ellas. De manera similar al caso de los modelos multinomiales multifactorizados, se pueden designar como parámetros de referencia a los parámetros asociados a la matriz de covarianzas de la primera factorización.



Independencia Condicional a través de la Matriz de Precisión: Sea  $X$  una variable aleatoria distribuida de forma normal y sea  $\{V, Y, Z\}$  una partición de  $X$ .

Sea  $W = \Sigma^{-1}$  la matriz de precisión del modelo, es decir, la inversa de la matriz de covarianzas  $\Sigma$ . Entonces, se cumple  $I(V, Y | Z)$  si y sólo si el bloque  $W_{VY}$  de la matriz  $W$  es la matriz nula. El teorema siguiente muestra que, para variables aleatorias normales, los términos dependencia y correlación son equivalentes, así como los términos dependencia condicional y correlación parcial.

## Modelos probabilísticos definidos Condicionalmente

En las secciones anteriores se han analizado los modelos multifactorizados, que permiten resolver el problema de compatibilidad de los modelos basados en multigrafos y los modelos basados en una lista de relaciones de independencia. En esta sección se trata el problema de la definición de un modelo probabilístico mediante un conjunto de funciones de probabilidad condicionada. Los modelos definidos de esta forma se denominan modelos probabilísticos definidos condicionalmente.



Modelos definidos condicionalmente: Considérese un conjunto de variables  $X = \{X_1, \dots, X_n\}$ . Un modelo probabilístico definido condicionalmente consiste en un conjunto de probabilidades marginales y condicionadas de la forma

$$P = \{p(u_i | v_i); i = 1, \dots, m\},$$

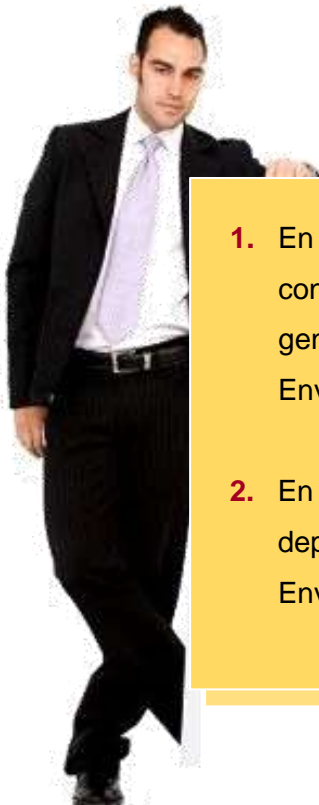
que define una única función de probabilidad de  $X$ , donde  $U_i$  y  $V_i$  son subconjuntos disjuntos de  $X$  y  $U_i \cap V_i = \emptyset$ .

# Lecturas Recomendadas

- ❖ **SISTEMAS EXPERTOS Y MODELOS DE REDES PROBABILÍSTICAS**  
<http://fismat.umich.mx/~htejeda/gutierjm/BookCGH.pdf>

- ❖ **SISTEMAS EXPERTOS PROBABILÍSTICOS**  
<http://es.slideshare.net/lcahuich/sistemas-expertosprobabilisticos-2a-parte>

## Actividades y Ejercicios



1. En un documento en Word realice un informe académico sobre la construcción de modelos gráficos y probabilísticos, haciendo énfasis en generación de gráficos no dirigidos.

Envíalo a través de "**Construcción de Modelos Gráficos**".

2. En un documento en Word presente dos ejemplos de modelos de dependencia: grafoides y semigrafoides.

Envíalo a través de "**Modelos de Dependencia**".

# Autoevaluación

- 1) Para comprobar si un grafo dirigido verifica una relación de independencia dada, es necesario introducir :
  - a. Un criterio de separación.
  - b. Tres grafos continuos.
  - c. Un grafo no dirigido.
  - d. Un grafo no continuo.
  - e. Dos o más uniones de grafos.
  
- 2) ¿Cuántos modelos para definir relaciones de independencia condicional hay?
  - a. 5.
  - b. 2.
  - c. 4
  - d. 7.
  - e. 3.
  
- 3) Es un conjunto de relaciones de independencia que es cerrado con respecto a las propiedades de simetría, descomposición, unión débil, contracción e intersección.
  - a. Un grafoide.
  - b. Inter-operatividad.
  - c. Exactitud.
  - d. Seguridad de acceso.
  - e. Fiabilidad.
  
- 4) Un modelo de dependencia se denomina probabilístico si contiene \_\_\_\_\_ dadas por una función de probabilidad conjunta.
  - a. Accesos.
  - b. Interoperatividad.
  - c. Todas las relaciones de independencia seguridad.
  - d. Exactitud.
  - e. Fiabilidad.
  
- 5) Los grafos son herramientas muy útiles para definir la estructura de \_\_\_\_\_ de un modelo probabilístico.
  - a. Discreción.
  - b. Indiscreción.
  - c. independencia.
  - d. Multi discreción.
  - e. Grafoide discreto.

- 6) Para representar relaciones de independencia condicional por medio de grafos no dirigidos se necesita definir de forma precisa:**
- a. Algoritmo y su aplicación.
  - b. Un criterio de separación apropiado.
  - c. Los caminos entre nodos.
  - d. Dependencia de los nodos.
  - e. Independencias asociadas de un grafo.
- 7) La estructura general de una función de probabilidad conjunta involucra un excesivo número de:**
- a. Códigos.
  - b. Variables.
  - c. Simplificaciones.
  - d. Parámetros.
  - e. Modelos.
- 8) Cada modelo probabilístico tiene asociado un modelo de:**
- a. Independencia.
  - b. Variables.
  - c. Parámetros.
  - d. Funciones.
  - e. Dependencia.
- 9) Las listas de independencias constituyen una alternativa a los modelos gráficos para la construcción de:**
- a. Modelos probabilísticos.
  - b. Relaciones.
  - c. Variables.
  - d. Factorización.
  - e. Independencia derivada.
- 10) ¿En qué modelo se analiza el problema de representar modelos probabilísticos para encontrar el grafo correspondiente a un modelo de dependencia probabilístico?**
- a. Modelos de dependencia gráficos no dirigidos.
  - b. Modelos a grafos no dirigidos.
  - c. Modelos de independencia gráficos no dirigidos.
  - d. Modelo de mapa de independencia.
  - e. Modelo de mapa de dependencia.

## UNIDAD DE APRENDIZAJE III: MODELOS PROBABILÍSTICOS Y GRÁFICOS

Los grafos son herramientas muy potentes para describir de forma intuitiva las relaciones de dependencia e independencia existentes en un conjunto de variables. Por tanto, una forma de definir un modelo probabilístico es partir de un grafo que describa las relaciones existentes entre las variables.

Para comprobar cuales, de entre todas las posibles relaciones de independencia condicional, son satisfechas por el grafo. Los criterios de separación gráfica son las reglas para entender cómo pueden codificarse dependencias e independencias en un grafo. Estos criterios dependen del tipo de grafo (dirigido o no dirigido) que se esté considerando.

En muchas situaciones prácticas, las relaciones existentes entre un conjunto de variables pueden ser representadas por un grafo no dirigido  $G$ . Cada variable puede ser representada por un nodo del grafo. Si dos variables son dependientes, esta relación puede representarse por un camino que conecte estos nodos. Por otra parte, si dos variables son independientes, entonces no deberá existir ningún camino que una estos nodos. De esta forma, el concepto de dependencia entre variables puede relacionarse con el concepto de conexión entre nodos.

Para comprobar si un grafo dirigido verifica una relación de independencia dada, es necesario introducir otro criterio de separación. Hasta ahora se han introducido tres modelos distintos para definir relaciones de independencia condicional: modelos probabilísticos, modelos gráficos no dirigidos, y modelos gráficos dirigidos.

[illegible]



# Introducción

## **a) Presentación y contextualización:**

Las redes Bayesianas son modelos gráficos probabilísticos utilizados en la toma de Decisiones. Una red Bayesiana representa una función de distribución conjunta Sobre un conjunto finito de variables. Muchas de las actividades en la ingeniería del software, como por ejemplo, la estimación de costes o esfuerzo, evaluación de riesgos o fiabilidad tratan con valores inciertos o probabilísticas. Por tanto, diversas técnicas estadísticas y la teoría de la probabilidad han sido aplicadas a la ingeniería del software desde sus inicios.

## **b) Competencia:**

**Explica la importancia del análisis y estudio de la propagación exacta en diversas redes probabilísticas.**

## **c) Capacidades:**

1. Comprende las generalidades y aplicación de la propagación de evidencias.
2. Conoce los principales métodos de propagación aproximada, identificando las características que la representan.
3. Reconoce la importancia de la propagación simbólica de evidencia respecto al desarrollo de redes.
4. Aplica las diversas teorías de aprendizaje sobre las redes bayesianas en diversos sistemas expertos.

## **d) Actitudes:**

- ✓ Muestra interés por el análisis sobre la propagación exacta en diversas redes probabilísticas.
- ✓ Muestra entusiasmo en los diversos desarrollos de las teorías respecto a la propagación en redes.

## **e) Presentación de Ideas básicas y contenidos esenciales de la Unidad:**

**La Unidad de Aprendizaje 04: Propagación Exacta en Redes Probabilísticas,** comprende el desarrollo de los siguientes temas:

**TEMA 01: Propagación de Evidencia.**

**TEMA 02: Métodos de Propagación Aproximada.**

**TEMA 03: Propagación Simbólica de Evidencia.**

**TEMA 04: Aprendizaje en Redes Bayesianas.**

# TEMA 1

## *Propagación de Evidencia*



### Competencia:

**Comprender las generalidades y aplicación de la propagación de evidencias.**



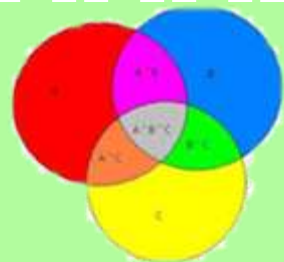
# Desarrollo de los Temas



## Tema 01: Propagación de Evidencia

La propagación de evidencia es una de las tareas más importantes de un sistema experto, pues permite obtener conclusiones cuando se dispone de nueva información (síntomas, etc.). Supóngase un conjunto de variables discretas  $X = \{X_1, \dots, X_n\}$  y una función de probabilidad  $p(x)$ , en  $X$ . Cuando no se dispone de ninguna información, es decir, cuando no existe evidencia, el proceso de propagación consiste en calcular las probabilidades marginales  $p(X_i = x_i)$ , también denotadas por  $p(x_i)$ , para cada  $X_i \in X$ . Estas probabilidades proporcionan información “a priori” sobre los distintos valores que pueden tomar las variables.

Cuando se dispone de cierta evidencia, es decir, cuando se conoce un conjunto de variables  $E \subset X$  que tienen asociadas los valores  $X_i = e_i$ , para  $X_i \in E$ , el proceso de propagación debe tener en cuenta estos valores para calcular las nuevas probabilidades de los nodos.



Evidencia. Un subconjunto de variables  $E \subset X$  cuyos valores son conocidos,  $E = e$ , en una situación dada, se conoce como conjunto de evidencia, o simplemente evidencia.



En esta situación, la propagación de evidencia consiste en calcular las funciones de probabilidad condicionada  $p(x_i | e)$  para cada variable  $X_i \in E$ , dada la evidencia  $E = e$ . Estas funciones de probabilidad condicionada miden el efecto producido por la evidencia en cada variable. Cuando no se dispone de evidencia ( $E = \emptyset$ ), las funciones condicionadas  $p(x_i | e)$  son simplemente las funciones de probabilidad marginal  $p(x_i)$ . Una forma de calcular las probabilidades  $p(x_i | e)$  consiste en utilizar la fórmula que implica  $p(x_i | e) = p(x_i, e) p(e) \propto p(x_i, e)$ , donde  $1/p(e)$  es una constante de proporcionalidad. Por tanto, se puede obtener  $p(x_i | e)$ , calculando y normalizando las probabilidades marginales  $p(x_i, e)$ .

De esta forma se tiene  $p(x_i, e) = \sum_{pe(x_1, \dots, x_n)} p_e(x_1, \dots, x_n)$ , donde  $p_e(x_1, \dots, x_n)$  es la función de probabilidad obtenida sustituyendo en  $p(x_1, \dots, x_n)$  las variables con evidencia,  $E$ , por sus valores  $e$ . Por tanto, para calcular  $p(x_i, e)$ , ha de sumarse  $p_e(x_1, \dots, x_n)$  para todas las posibles combinaciones de valores de las variables que no estén contenidas en  $E$ , excepto la variable  $X_i$ .

Debido al elevado número de combinaciones de valores que involucra, este método de “fuerza bruta” resulta altamente ineficiente, incluso en redes con un número reducido de variables. Por ejemplo, en el caso de variables binarias, la ecuación requiere la suma de  $2^{n-1}$  probabilidades distintas. En la Figura 8.1 se muestra el tiempo de computación necesario para calcular  $p(x_i)$  en un ordenador personal. Esta figura muestra que el tiempo de computación crece de forma exponencial con el número de variables del modelo,  $n$ . Puede observarse que este método es ineficiente incluso para modelos con solo unas decenas de variables.

## PROPAGACIÓN EN POLIÁRBOLES

El poliarbol es uno de los modelos gráficos más simples para construir redes Bayesianas. La característica principal de este algoritmo es que su complejidad es lineal en el tamaño de la red (es decir en el número de nodos y aristas que la componen), a diferencia del método de fuerza bruta que requiere un número exponencial de operaciones para realizar la propagación.



Por ejemplo, el nodo  $D$  divide al poliarbol en dos poliarboles inconexos, el primero de los cuales,  $\{A, B, C\}$ , incluye a sus padres y a los nodos que son accesibles desde  $D$  a través de sus padres, y el segundo,  $\{E, F, G\}$ , que incluye a sus hijos y a los nodos que son accesibles desde  $D$  a través de sus hijos. en la cual también puede comprobarse que el nodo  $D$  separa a estos dos conjuntos, es decir, que se verifica graficamente la relación de independencia  $I(\{A, B, C\}, \{E, F, G\} | D)$ .

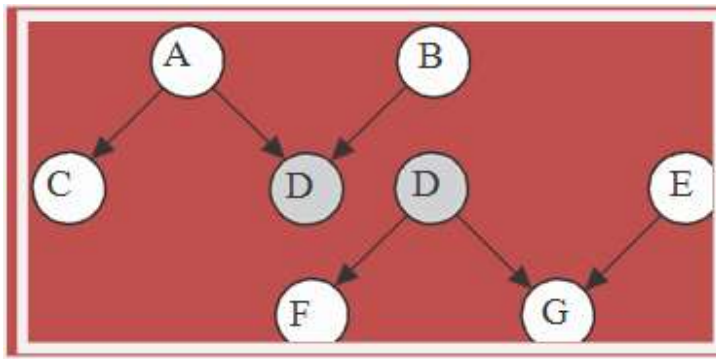
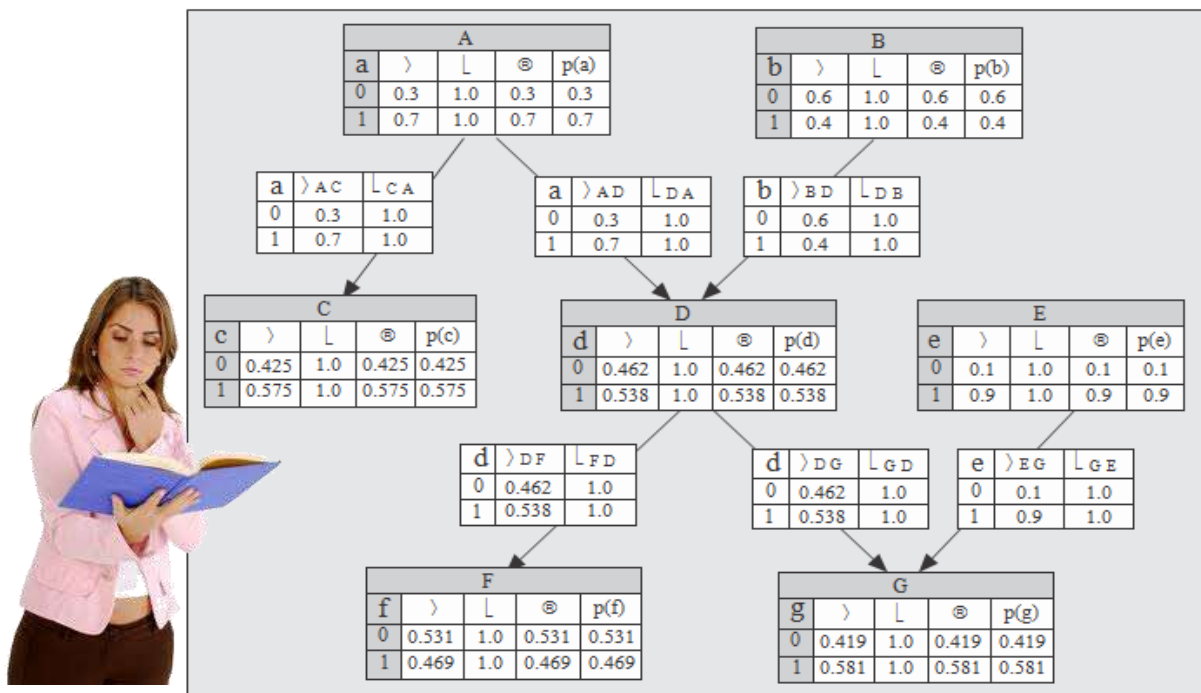


Figura 8.1. El nodo D divide al polígrafo en dos polígrafos inconexos.

El proceso de propagación puede realizarse en este tipo de grafos de un modo eficiente combinando la información procedente de los distintos subgrafos mediante el envío de mensajes (cálculos locales) de un subgrafo a otro.



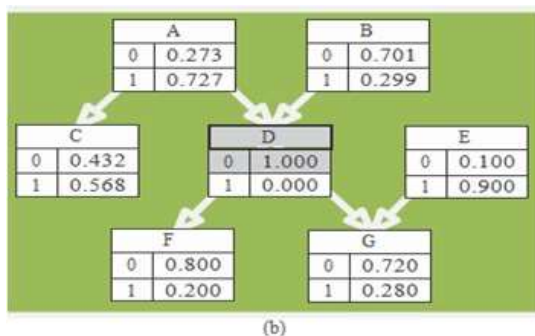
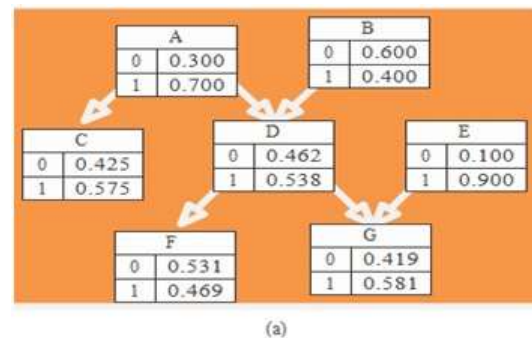
Valores numéricos de los mensajes y funciones calculados por el algoritmo de propagación en polígrafos cuando no se dispone de evidencia.



## PROPAGACIÓN EN REDES MÚLTIPLEMENTE CONEXAS

El método de propagación en poliárboles descrito en la sección anterior es válido solamente para redes de estructura simple (poliárboles), en las cuales existe un único camino entre cada par de nodos. Por tanto, este tipo de redes carecen de generalidad y no son aplicables en numerosas situaciones prácticas. En estos casos es necesario trabajar con grafos múltiplemente conexos (grafos que contienen bucles) en los que pueden existir varios caminos entre dos nodos. Dos de los métodos de propagación más importantes para este tipo de redes son los denominados métodos de condicionamiento y método de agrupamiento.

La idea fundamental del método de propagación por condicionamiento es cortar los múltiples caminos entre los nodos mediante la asignación de valores a un conjunto reducido de variables contenidas en los bucles. De esta forma se tendrá un poliárbol en el cual se podrá aplicar el algoritmo de propagación para poliárboles descrito en la sección anterior. Por otra parte, el método de agrupamiento construye representaciones auxiliares, de estructura más simple, uniando conjuntos de nodos del grafo original (por ejemplo, un árbol de unión). De esta forma se puede obtener un grafo con estructura de poliárbol en el que pueden aplicarse las mismas ideas descritas en la sección anterior para propagar evidencia.



**Probabilidades marginales (iniciales) de los nodos (a) y probabilidades condicionadas (actualizadas), dada la evidencia  $D = 0$  (b).**



### MÉTODO DE CONDICIONAMIENTO

En el caso de redes Bayesianas múltiplemente conexas ya no se cumple la propiedad de que un nodo cualquiera separa el grafo en dos partes inconexas. Por tanto, algunas de las propiedades de independencia aplicadas en el algoritmo de propagación en poliárboles no pueden ser aplicadas en esta situación.

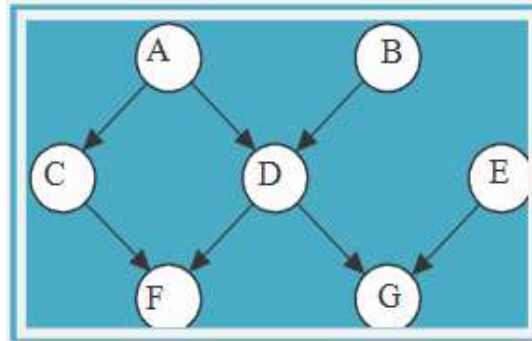


FIGURA 8.16. Grafo múltiplemente conexo

La idea básica del algoritmo de condicionamiento es cortar estas vías alternativas de comunicación contenidas en los bucles asignando un valor arbitrario a un conjunto de nodos. Este conjunto de nodos se suele denominar conjunto de corte (en inglés, cutset). Por ejemplo, el nodo D no separa al grafo de la Figura en dos partes inconexas, pero si se considera el conjunto de corte formado por el nodo C, entonces, el conjunto {C, D} separa a {A, B} de {E, F, G}, los subgrafos que contienen a los padres e hijos de D, respectivamente. Por tanto, se puede cortar el bucle contenido en el grafo considerando el nodo C como un nodo evidencial, es decir, asignándole un valor arbitrario.

Esta idea de cortar los bucles para obtener un grafo de estructura más simple puede ser llevada a la práctica utilizando el método denominado absorción de evidencia. Este método muestra que la evidencia puede ser absorbida por el grafo cambiando su topología. De forma más precisa, si  $X_i$  es un nodo evidencial, se pueden eliminar del grafo todas las aristas de la forma  $X_i \rightarrow X_j$  sustituyendo la función de probabilidad condicionada del nodo  $X_j$ ,  $p(x_j | \pi_j)$ , por una función definida sobre un conjunto más reducido de variables:

$$p_1(x_j | \pi_j \setminus x_i) = p(x_j | \pi_j \setminus x_i, X_i = e_i).$$

Esta operación deja inalterado el modelo probabilístico, mientras que implica la topología del grafo al eliminar un conjunto de aristas. Obsérvese que el conjunto  $\Pi_j \setminus X_i$  es el nuevo conjunto de padres del nodo  $X_j$  en el grafo modificado.



Por ejemplo, si se asigna un valor arbitrario,  $C = c$ , al nodo  $C$ , es decir, si se convierte  $C$  en un nodo evidencial en el grafo de la Figura 8.16, entonces se puede absorber esta evidencia eliminando del grafo la arista  $C \rightarrow F$ , obteniendo así un nuevo grafo con estructura de poliarbol (ver Figura 8.17).

Para mantener inalterada la función de probabilidad condicionada del conjunto de variables no evidenciales,  $p(y|C = c)$ , se reemplaza la función de probabilidad  $p(f|c, d)$  por  $p_1(f|d) = p(f|C = c, d)$ , lo cual elimina la dependencia del nodo  $F$  respecto de la evidencia  $C$ .

$$p_1(f|d) = p(f|C=c, d)$$

Absorción de la evidencia  $C = c$  mediante la arista  $C \rightarrow F$ .

Por tanto, utilizando el método de absorción de evidencia se puede reducir un grafo múltiplemente conexo a un poliárbol, asignando un valor arbitrario a los nodos de un conjunto de corte  $C = \{C_1, \dots, C_m\}$ .



### MÉTODOS DE AGRUPAMIENTO

El algoritmo de propagación en poliárboles y el algoritmo de condicionamiento introducidos en las secciones anteriores aprovechan la estructura particular de los grafos dirigidos para propagar la evidencia. Por tanto, estos algoritmos son sólo aplicables a redes Bayesianas. En esta sección se presenta un método de propagación distinto, el método de agrupamiento que, a partir de las estructuras locales contenidas en el grafo, produce representaciones alternativas para propagar la evidencia. Por tanto, estos métodos no dependen del tipo de grafo y son aplicables tanto a redes de Markov, como a redes Bayesianas.

El método de agrupamiento, inicialmente desarrollado por Lauritzen y Spiegelhalter (1988), se basa en la construcción de subconjuntos de nodos (aglomerados) que capturen las estructuras locales del modelo probabilístico asociado al grafo. De esta forma, el proceso de propagación de evidencia puede ser realizado calculando probabilidades locales (que dependen de un número reducido de variables), evitando así calcular probabilidades globales (que dependen de todas las variables), los conglomerados de un grafo son los subconjuntos que representan sus estructuras locales.

Por tanto, en primer lugar, el algoritmo de agrupamiento calcula los conglomerados del grafo; a continuación obtiene las funciones de probabilidad condicionada de cada conglomerado calculando de forma iterativa varias funciones de probabilidad locales. Por último, se obtiene la función de probabilidad condicionada de cada nodo marginalizando la función de probabilidad de cualquier conglomerado en el que esté contenido. En esta sección se presentan dos versiones de este algoritmo, una para redes de Markov y otra para redes Bayesianas.

**Eliminar de X los nodos evidenciales.** Este proceso también implica modificar el conjunto de conglomerados y la representación potencial. La nueva representación potencial,  $(C^*, \Psi^*)$ , está definida en  $X^*$ , donde  $X^* = X \setminus E$ ,  $C^*$  es el nuevo conjunto de conglomerados y  $\Psi^*$  son los nuevos potenciales, que contienen la evidencia, y que han sido obtenidos de la forma siguiente: Para cada conglomerado  $C_i$  contenido en  $C$  tal que  $C_i \cap E = \emptyset$ , se incluye el conjunto  $C_i \setminus E$  en  $C^*$  y se define Para el resto de los conglomerados que no contienen nodos evidenciales, no es necesario realizar ninguna modificación en las representaciones potenciales correspondientes. Con ello, se tiene  $p(x^*|e) \propto \prod_{i=1} \psi^*(c_i)$ .

Por tanto, en ambos casos, se puede aplicar el método anterior para obtener la función de probabilidad condicionada de los nodos, dada la evidencia  $E = e$ . En el primer caso se continua con la misma estructura utilizan más recursos de los necesarios. En el segundo caso, no se utilizan más recursos de los necesarios, pero se necesita modificar la estructura. Por tanto, se requiere un consenso entre ambas opciones con objeto de elegir la más adecuada en cada caso.

### Algoritmo de Agrupamiento en Redes Bayesianas

En la sección anterior se presentó el método de agrupamiento para propagar evidencia en redes de Markov. En esta sección se presenta una adaptación

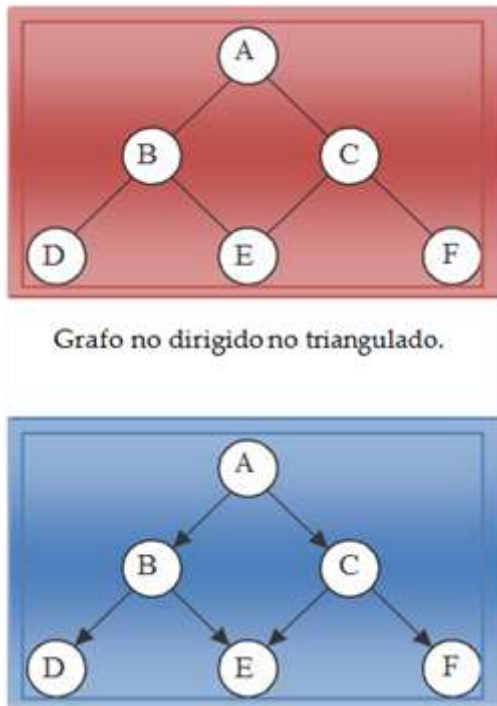


FIGURA 8.26. Grafo dirigido acíclico múltiplemente conexo.

### PROPAGACIÓN EN ARBOLES DE CONGLOMERADOS

El algoritmo de agrupamiento agrupa conjuntos de nodos con cierta estructura local creando una cadena de conglomerados para propagar evidencia. Algunas modificaciones de este método utilizan una representación gráfica de la cadena de conglomerados (por ejemplo, un árbol de unión) para propagar la evidencia de forma más eficiente. El método de los universos de conocimiento desarrollado por Jensen, Olesen y Andersen Transforma el grafo múltiplemente conexo en un árbol de conglomerados asociado al grafo original.

# *Métodos de Propagación Aproximada*

## TEMA 2



### Competencia:

**Conocer los principales métodos de propagación aproximada, identificando las características que la representan.**







## Tema 02: Métodos de Propagación Aproximada

### BASE INTUITIVA DE LOS MÉTODOS DE SIMULACIÓN

En esta sección se ilustra un esquema general de simulación mediante un sencillo ejemplo. Considérese una urna que contiene seis bolas numeradas  $\{1, \dots, 6\}$ . Supóngase que se quiere realizar el siguiente experimento. Se selecciona una bola al azar de la urna, se apunta su número, se devuelve a la urna, y se mezclan las bolas antes de proceder a extraer la bola siguiente. Este esquema de muestreo se denomina muestreo con reemplazamiento. Cada selección de una bola se llama una extracción o un experimento. En este caso cada extracción tiene seis posibles resultados,  $\{1, \dots, 6\}$ .

Sea  $X_i$  el resultado (el número de la bola) de la extracción  $i$ -ésima.



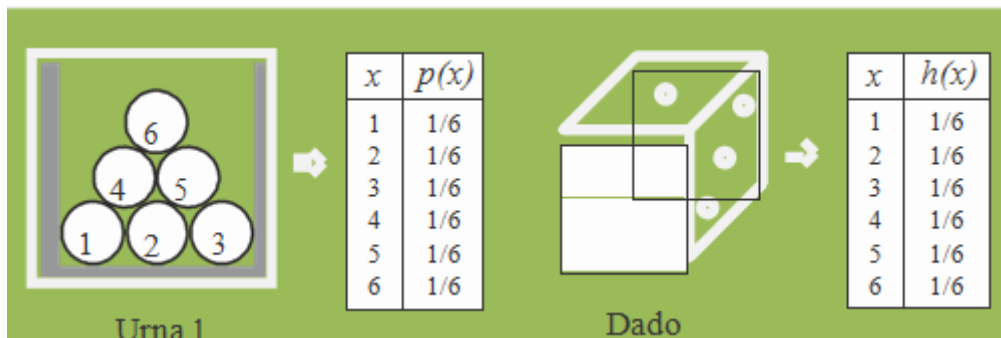
Puesto que el muestreo se hace con reemplazamiento, las extracciones son independientes (el resultado de una extracción no influye en el resultado de las demás). Claramente,  $X_i$  es una variable uniforme con función de probabilidad  $p(X_i = x_i) = 1/6$ , para  $x_i = 1, \dots, 6$  y  $i = 1, \dots, N$ , donde  $N$  es el número de extracciones (el tamaño de la muestra). Utilizando esta función de probabilidad conjunta, se pueden calcular las probabilidades exactas de ciertos sucesos tales como  $p(X_1 = 1, \dots, X_n = 1)$   $p$  (número de pares = número de impares), etc.

Estos cálculos son fáciles en este caso puesto que la distribución es uniforme (hay exactamente una bola para cada uno de los números  $\{1, \dots, 6\}$ ), las extracciones son idénticas (se usa la misma urna), y el resultado de cada extracción es independiente de los resultados de los demás (muestreamos con reemplazamiento). Los cálculos de las probabilidades exactas son complicados y costosos cuando la distribución no es uniforme (por ejemplo, se tiene distinto número de bolas de diferentes tipos), las extracciones no son idénticas (por ejemplo, se realiza un muestreo con diferentes números de bolas), y/o extracciones que no son independientes (por ejemplo, muestreo sin reemplazamiento).





En estas situaciones complicadas, se pueden calcular las probabilidades de ciertos sucesos de forma aproximada mediante técnicas de simulación. Se puede, por ejemplo, repetir un experimento  $N$  veces. Se obtiene lo que se llama una muestra de tamaño  $N$ . Entonces, la probabilidad de un suceso puede aproximarse por el cociente entre el número de veces que ocurre dicho suceso y el número total de simulaciones  $N$ . Claramente, cuanto mayor es el tamaño de la muestra más aproximada será la aproximación.



**Simulando la extracción de bolas con reemplazamiento de la Urna y mediante un dado.**

Que es más fácil lanzar el dado que extraer la bola de una urna, devolverla y mezclar las bolas antes de la extracción siguiente. En otras palabras, si no es fácil obtener muestras de la distribución de la población se debe elegir otra distribución que resulte más sencilla para la simulación. se puede utilizar un dado para simular la extracción de bolas de urnas con diferentes números de bolas? La respuesta, afortunadamente, es positiva. Por ejemplo, supóngase que la urna contiene solo cinco bolas numeradas  $\{1, \dots, 5\}$  (Urna 2). Sea  $X$  el número de bolas con el número  $i$  sacadas al azar con reemplazamiento de la Urna 2.

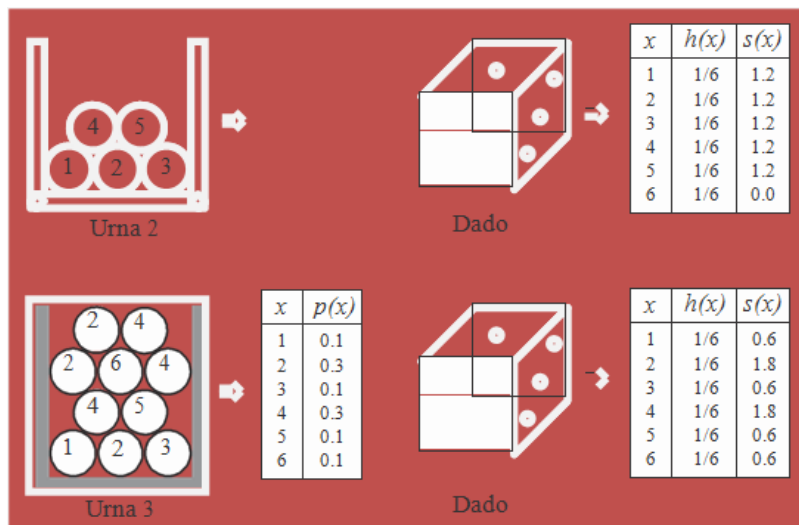
Entonces  $X$  es una variable aleatoria cuya función de probabilidad,  $p(x)$ , se muestra en la Figura 9.2 (Urna 2). En este caso, la distribución simulada (el dado) no es la misma que la distribución de la población (Urna 2), es decir,  $p(x) \neq h(x)$  (las columnas etiquetadas  $s(x)$  se explicaran en breve). A pesar del hecho de que la Urna 2 y el dado no tienen la misma distribución, se puede todavía utilizar el dado para simular la extracción de bolas de la Urna 2, pero se tiene que corregir por el hecho de que las distribuciones de la población y la simulada no coinciden.

Una forma de tener en cuenta esta diferencia es la siguiente: cuando en el dado sale un 6, se ignora la tirada y se repite de nuevo hasta que salga un valor menor que 6, en cuyo caso se hace y igual al número que salga y se toma y como valor generado de la población  $p(x)$ . Este ejemplo es en realidad un caso especial del método conocido como método de aceptación- rechazo.

**El método de aceptación - rechazo.** Sea  $X$  una variable aleatoria con función de probabilidad  $p(x)$ . Supóngase que  $p(x)$  puede ser expresada como

$$p(x) = c g(x) h(x), \quad (9.2)$$

donde  $c \geq 1$ ,  $0 \leq g(x) \leq 1$  y  $h(x)$  es una función de probabilidad. Sea  $U$  una variable aleatoria uniforme  $U(0, 1)$  y sea  $Y$  una variable aleatoria con función de probabilidad  $h(y)$  independiente de  $U$ . Entonces, la función de probabilidad condicional de  $Y$  dado que  $u \leq g(y)$  coincide con la función de probabilidad de  $X$ . Por otra parte, la probabilidad de aceptar la muestra (eficiencia) es  $1/c$ .



Una ilustración de un esquema general de simulación.

Por ejemplo, en el caso de la Urna 2 que se muestra en la Figura 9.2, se puede escribir  $p(x) = cg(x)h(x)$ , donde  $p(x)$  y  $h(x)$  se muestran en la Figura 9.2,  $c = 6/5$  y 0, si  $x = 6$ ,  $g(x) =$

Por ello, utilizando el teorema anterior, se puede obtener una muestra de  $p(x)$  (Urna 2) usando  $h(x)$  (el dado) y comprobando la condición  $u \leq g(x)$  para todo valor  $x$  que se simule de  $h(x)$ , donde  $u$  es un número obtenido de la distribución uniforme  $U(0, 1)$ . Por tanto, en este caso, el suceso  $x = 6$  siempre se rechaza, ya que  $g(6) = 0$ , y los restantes sucesos se aceptan siempre.

# *Propagación Simbólica de Evidencia*

## TEMA 3



### Competencia:

**Reconocer la importancia de la propagación simbólica de evidencia respecto al desarrollo de redes.**





## Tema 03: Propagación Simbólica de Evidencia

### NOTACIÓN Y CONCEPTOS PRELIMINARES

Se ha visto que la función de probabilidad conjunta asociada a las redes probabilísticas de Markov descomponibles y Bayesianas puede darse mediante una factorización como producto de probabilidades condicionales

$$p(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i | \pi_i).$$

En el caso de redes Bayesianas, los conjuntos condicionantes son los padres del nodo,  $\Pi_i$ ,  $i = 1, \dots, n$ . En el caso de redes de Markov descomponibles, estos conjuntos se obtienen aplicando la regla de la cadena a la factorización obtenida a partir de la cadena de conglomerados. Por tanto, aunque algunos de los métodos introducidos en este capítulo pueden ser fácilmente extendidos para tratar una representación potencial de la de probabilidad conjunta, por simplicidad, pero sin pérdida de generalidad, se utiliza el conjunto de probabilidades condicionales en como representación paramétrica básica de la función de probabilidad conjunta.



Sea  $X = \{X_1, \dots, X_n\}$  un conjunto de  $n$  variables discretas, cada una de las cuales puede tomar valores en el conjunto  $\{0, 1, \dots, r_i\}$ , y sea  $B = (D, P)$  una red Bayesiana definida sobre  $X$ , donde el grafo dirigido acíclico  $D$  determina la estructura del conjunto de probabilidades condicionales, y  $P = \{p(x_1 | \pi_1), \dots, p(x_n | \pi_n)\}$  es el conjunto de probabilidades condicionales que se necesitan para especificar la función de probabilidad conjunta. Algunas de las probabilidades condicionales en (10.1) pueden darse en forma numérica y otras en forma simbólica, es decir,  $p(x_i | \pi_i)$  pueden ser familias paramétricas o probabilidades totalmente especificadas numéricamente.

Nodo Simbólico. Cuando  $p(x_i | \pi_i)$  es una familia paramétrica simbólica (es decir, depende de al menos un parámetro en forma simbólica), el nodo  $X_i$  se denomina un nodo simbólico, y se utiliza  $\Theta_i$  para denotar sus correspondientes parámetros simbólicos. Cuando  $p(x_i | \pi_i)$  es una familia paramétrica, es decir, cuando  $X_i$  es un nodo simbólico, una elección conveniente de los parámetros es la siguiente

$$\theta_{ij\pi} = p(X_i = j | \Pi_i = \pi), \quad j \in \{0, \dots, r_i\}$$

Donde  $\pi$  es cualquier posible realización de los padres,  $\Pi_i$ , de  $X_i$ . Por ello, el primer subíndice de  $\theta_{ij\pi}$  se refiere al número del nodo, el segundo subíndice se refiere al estado del nodo, y los restantes subíndices se refieren a las realizaciones de sus padres. Puesto que  $\sum_{j=0}^{r_i} \theta_{ij\pi} = 1$ , para todo  $i$  y  $\pi$ ,

No todos los parámetros son libres, es decir, uno cualquiera de ellos puede ser escrito como la unidad menos la suma del resto. Por ejemplo, el primer parámetro puede escribirse como

$$\theta_{i0\pi} = 1 - \sum_{j=1}^{r_i} \theta_{ij\pi}.$$



Para implicar la notación en los casos en los que la variable  $X_i$  no tiene padres, se utiliza  $\theta_{ij}$  para denotar  $p_i(X_i = j)$ ,  $j \in \{0, \dots, r_i\}$ . Se ilustra esta notación usando el ejemplo siguiente.

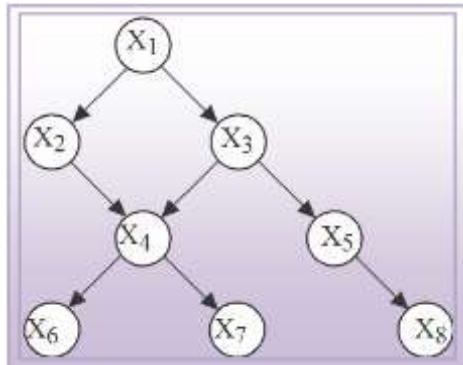
Ejemplo de Nodos simbólicos. Considérese una red Bayesiana discreta consistente en las variables  $X = \{X_1, \dots, X_8\}$ . La estructura del grafo implica que la probabilidad conjunta del conjunto de nodos puede escribirse en la forma

$$p(x) = p(x_1) p(x_2 | x_1) p(x_3 | x_1) p(x_4 | x_2, x_3) p(x_5 | x_3) p(x_6 | x_4) p(x_7 | x_4) p(x_8 | x_5).$$

Por simplicidad, y sin pérdida de generalidad, supóngase que todos los nodos representan variables binarias con valores en el conjunto  $\{0, 1\}$ . Esto y la estructura de la distribución de probabilidad implica que la función de probabilidad conjunta de las ocho variables depende de 34 parámetros  $\Theta = \{\theta_{ij\pi}\}$ .



Nótese, sin embargo, que solamente 17 de ellos son libres (puesto que las probabilidades en cada una de las probabilidades condicionales deben sumar la unidad). Estos 17 parámetros se dan en la Tabla.



Un grafo dirigido acíclico.



$X$	$\Pi$	Parametros libres
$X_1$	$\varphi$	$\theta_{10} = p(X_1 = 0) = 0.2$
$X_2$	$X_1$	$\theta_{200} = p(X_2 = 0/X_1 = 0) = 0.3$ $\theta_{201} = p(X_2 = 0/X_1 = 1) = 0.5$
$X_3$	$X_1$	$\theta_{300} = p(X_3 = 0/X_1 = 0)$ $\theta_{301} = p(X_3 = 0/X_1 = 1) = 0.5$
$X_4$	$X_2, X_3$	$\theta_{4000} = p(X_4 = 0/X_2 = 0, X_3 = 0) = 0.1$ $\theta_{4001} = p(X_4 = 0/X_2 = 0, X_3 = 1) = 0.8$ $\theta_{4010} = p(X_4 = 0/X_2 = 1, X_3 = 0) = 0.3$ $\theta_{4011} = p(X_4 = 0/X_2 = 1, X_3 = 1) = 0.4$
$X_5$	$X_3$	$\theta_{500} = p(X_5 = 0/X_3 = 0) = 0.3$ $\theta_{501} = p(X_5 = 0/X_3 = 1) = 0.1$
$X_6$	$X_4$	$\theta_{600} = p(X_6 = 0/X_4 = 0)$ $\theta_{601} = p(X_6 = 0/X_4 = 1) = 0.9$
$X_7$	$X_4$	$\theta_{700} = p(X_7 = 0/X_4 = 0) = 0.3$ $\theta_{701} = p(X_7 = 0/X_4 = 1) = 0.6$
$X_8$	$X_5$	$\theta_{800} = p(X_8 = 0/X_5 = 0) = 0.2$ $\theta_{801} = p(X_8 = 0/X_5 = 1) = 0.4$

TABLA de El conjunto de parámetros libres asociados a las distribuciones condicionales.



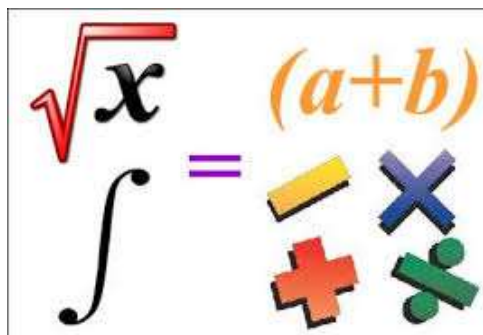
En este ejemplo, solo los nodos X3 y X6 son nodos simbólicos puesto que sus correspondientes funciones de probabilidad condicionada contienen al menos un parámetro simbólico. Se tienen los conjuntos de parámetros  $\Theta_3 = \{\theta_{300}, \theta_{310}\}$  y  $\Theta_6 = \{\theta_{600}, \theta_{610}\}$ . Nótese que estos conjuntos incluyen todos los parámetros simbólicos, no sólo los parámetros libres. Por ello, el conjunto de parámetros simbólicos asociados a la red Bayesiana es  $\Theta = \{\Theta_3, \Theta_6\}$ .

### GENERACIÓN AUTOMÁTICA DE CÓDIGO SIMBÓLICO



El tratamiento con parámetros simbólicos es idéntico al tratamiento con valores numéricos, con la única diferencia de que las operaciones requeridas deben realizarse con un programa capaz de manipular símbolos en vez de números. Los cálculos simbólicos, sin embargo, son mucho más lentos que los numéricos y requieren más memoria.

Sin embargo, este método de resolver el problema es muy costoso computacionalmente, y resulta ineficiente incluso con números reducidos de variables. Una alternativa a este método consiste en adaptar algunos de los algoritmos de propagación numérica muestran que la adaptación simbólica de estos métodos requiere solo pequeñas modificaciones. Por ejemplo, el algoritmo de propagación por agrupamiento puede adaptarse fácilmente a la propagación simbólica utilizando una herramienta informática simbólica, tal como Matemática.



# *Aprendizaje en Redes Bayesianas*

## TEMA 4



### Competencia:

**Aplicar las diversas teorías de aprendizaje sobre las redes bayesianas en diversos sistemas expertos.**





## Tema 04: Aprendizaje en Redes Bayesianas

### MIDIENDO LA CALIDAD DE UNA RED BAYESIANA

Una medida de calidad,  $Q(B|S, \xi)$ , es un criterio mediante el cual se puede ordenar el conjunto de todas las redes Bayesianas posibles por su calidad, donde  $B$  es una red Bayesiana,  $\xi$  la información "a priori", y  $S$  un conjunto de datos. Por ello, dada la información "a priori"  $\xi$  y/o un conjunto de datos  $S$ , nuestro objetivo consiste en obtener una red Bayesiana de alta calidad. Una medida de calidad debe satisfacer algunas propiedades deseables. Por ejemplo, debe asignarse la misma calidad a las redes que conduzcan a la misma estructura de independencia.

A continuación se define esta importante propiedad.

### Equivalencia en Peso:

Dado un conjunto de datos  $S$ , una medida de calidad  $Q(B|S, \xi)$  se dice que es equivalente en peso si asigna el mismo valor a todo par de redes Bayesianas equivalentes  $B_1$  y  $B_2$ , es decir, si  $Q(B_1|S, \xi) = Q(B_2|S, \xi)$ .

Otras propiedades de las medidas de calidad son:

- A las redes recomendadas por los expertos se les debe asignar calidades más altas que a las rechazadas por ellos.
- Las representaciones perfectas deben recibir calidades mayores que las imperfectas.
- Las I -representaciones mínimas deben recibir calidades mayores que las no mínimas.

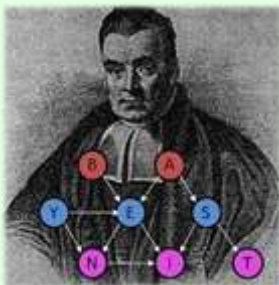
- Las redes con reducido número de parámetros a igualdad del resto de propiedades deben recibir calidades mayores que las de elevado número de parámetros.
- A las redes que confirmen la información contenida en los datos debe asignársele una calidad mayor que a aquellas que contradigan a estos.



Para ampliar conocimientos sobre estas y otras propiedades se remite al lector a consultar el trabajo de Bouckaert (1995).

Las medidas de calidad dependen de la incertidumbre de la información disponible. Dos posibles situaciones son:

1. Una situación en la que las estructuras probabilísticas y gráficas están ambas sometidas a incertidumbre. En este caso, se dispone de la información “a priori”  $\xi$  y el conjunto de datos  $S$ , y el objetivo consiste en encontrar la mejor red Bayesiana  $B(\theta) = (D, P(\theta))$  usando algún criterio de calidad. Nótese que  $\xi$  contiene información “a priori” referente a ambas estructuras, la gráfica y la paramétrica. Dados  $\xi$  y  $S$ , la calidad de una red Bayesiana  $B(\theta)$  depende de la calidad de sus subcomponentes,  $D$  y  $P(\theta)$ . Se usa:  
 $Q(B(\theta)|S, \xi)$  o  $Q(D, P(\theta)|S, \xi)$ .



Para denotar la medida de calidad de la red Bayesiana en su totalidad y para indicar que la medida depende de  $S$  y  $\xi$ . Sin embargo, en algunos casos se puede estar interesado sólo en el aprendizaje estructural. En tales casos, se puede obtener una medida de la calidad de la estructura gráfica maximizando la calidad de sus redes

Bayesianas  $Q(B(\theta)|S, \xi)$  con respecto a  $\theta$ , es decir,

$Q(D|S, \xi) = Q(D, P(\hat{\theta})|S, \xi)$ , Donde  $\hat{\theta}$  es el valor de  $\theta$  que maximiza  $Q(D, P(\theta)|S, \xi)$ . Alternativamente, se puede usar cualquier otra estimación de  $\theta$ , tal como la estimación de máxima verosimilitud, una estimación Bayesiana, etc.

2. Una situación en la que la estructura gráfica  $D$  es conocida y sólo la estructura probabilística está sometida a incertidumbre. En este caso, se está interesado sólo en el aprendizaje paramétrico, y el objetivo consiste en encontrar



la mejor estructura probabilística  $P(\theta)$ , utilizando algún criterio de calidad. Dados  $S$ ,  $D$  y  $\xi$ , la calidad de  $P(\theta)$  depende de la calidad de los parámetros estimados. Se usa  $Q(P(\theta)|D, S, \xi)$  para denotar la medida de calidad de la estructura probabilística de la red Bayesiana y para enfatizar que esta es condicionada a  $D$ ,  $S$ , y  $\xi$ . Nótese que  $\xi$  sólo contiene información “a priori” sobre la estructura paramétrica ya que se conoce con certeza la estructura tráfico.



Algunas medidas de calidad se definen como la suma de tres términos o componentes:  $Q = f(\text{información "a priori"}) + g(\text{datos disponibles}) + h(\text{complejidad})$ , donde  $f(\cdot)$ ,  $g(\cdot)$  y  $h(\cdot)$  Son funciones conocidas. El significado de estos términos se explica a continuación:

1. **La información "a priori":** La función  $f$  (información "a priori") asigna una probabilidad alta a las redes que han sido indicadas como altamente probables por la información "a priori" y una probabilidad baja a las que han sido indicadas como poco probables. Cuanto mayor sea la contribución de este término a la medida de calidad, mayor será el peso del conocimiento "a priori" frente al aportado por los datos. Este término contribuye decisivamente a la calidad de la red en estudio cuando no existen datos disponibles o son muy reducidos, pero es despreciable cuando los datos disponibles son abundantes.

Una elección típica para este término es  $\log p(B)$ , donde  $p(B) = p(D, \theta)$  es la probabilidad "a priori" asignada a la red  $B$ , donde  $\theta$  se usa en vez de  $P$  para mostrar la dependencia explícita de  $P$  del parámetro  $\theta$ . Si no hay conocimiento "a priori" disponible, este término se sustituye por cero, lo que es equivalente a suponer que  $p(B)$  es una distribución uniforme.



2. **Los datos disponibles:** La función  $g(\text{datos disponibles})$  es un término de bondad de ajuste que mide lo bien o mal que una red Bayesiana reproduce los datos  $S$ . Da una alta calidad a las redes que están de acuerdo con los datos y una baja calidad a las que los contradicen. La contribución de este término aumenta cuando se añaden aristas a la red. En tal caso se tienen más parámetros o grados de libertad y, normalmente, se puede obtener un mejor ajuste a los datos.

Algunas elecciones típicas para este término son las siguientes:

- (a) El logaritmo de la verosimilitud de los datos:  $\log p(S|D, \theta)$ .
- (b) El logaritmo de la probabilidad "a posteriori" de  $\theta$  dada la estructura  $D$  y los datos  $S$ :  $\log p(\theta | S, D)$ .

3. **La complejidad:** La función  $h$  (complejidad) penaliza las redes con estructura compleja (por ejemplo, redes con un gran número de aristas y/o un número alto de parámetros). Por ello, la función  $h()$  conduce a una calidad alta para las redes simples con un número reducido de aristas y parámetros, y a una baja calidad para las redes con muchas aristas y/o parámetros.
- Para medir la complejidad de una red Bayesiana es importante conocer su dimensión. Dimensión de una red Bayesiana. Sea  $X$  un conjunto de variables y  $B = (D, P)$  una red Bayesiana definida sobre  $X$ . La dimensión de esta red Bayesiana,  $\text{Dim}(B)$ , se define como el número de parámetros necesarios para especificar su función de probabilidad conjunta asociada.

Chickering (1995a) muestra que las redes Bayesianas independientemente equivalentes tienen la misma dimensión.

En la literatura existente se han propuesto varias medidas de calidad para redes Bayesianas. Estas se han clasificado en los tipos siguientes:

- Medidas de calidad Bayesianas.
- Medidas de mínima longitud de descripción.
- Medidas de información.



Estos tipos de medidas de calidad se discuten en las secciones siguientes.

**Medidas de Calidad Bayesianas.** En la teoría estadística Bayesiana, se supone inicialmente que la distribución “a priori”  $p(B) = p(D, \theta)$  la dan los expertos. Esta distribución refleja la opinión de los expertos sobre la frecuencia relativa de ocurrencia de diferentes redes Bayesianas  $B = (D, P(\theta))$ . Para mejorar el conocimiento, se obtienen unos datos  $S$  y, mediante el teorema de Bayes, la distribución “a posteriori”  $p(B, \theta|S)$  como sigue:

$$p(D, \theta|S) = \frac{p(D, \theta, S)}{p(S)} = \frac{p(D, \theta, S)}{\sum_{D, \theta} p(D, \theta, S)}$$



# Lecturas Recomendadas

❖ **CURSO BÁSICO DE SISTEMAS EXPERTOS**

<http://luisguillermo.com/cbse.pdf>

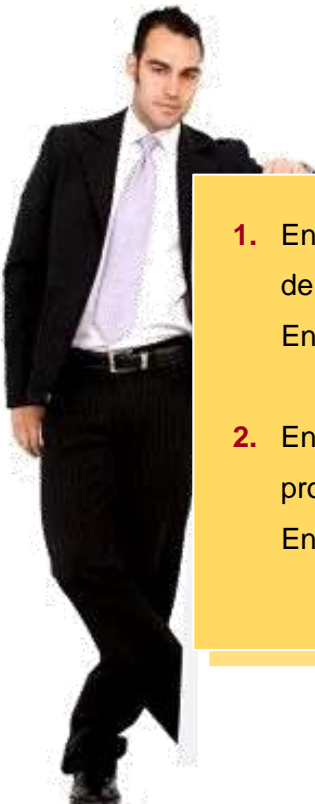
❖ **SISTEMAS EXPERTOS Y MODELOS DE REDES PROBABILÍSTICAS**

<http://fismat.umich.mx/~htejeda/gutierjm/BookCGH.pdf>

❖ **CONSTRUCCIÓN DE MODELOS PROBABILÍSTICOS**

[http://www.cs.us.es/cursos/iic-2010/Archivos/IIC%20-%20Teoria10\\_v03.pdf](http://www.cs.us.es/cursos/iic-2010/Archivos/IIC%20-%20Teoria10_v03.pdf)

## Actividades y Ejercicios



1. En un documento en Word presente dos ejemplos sobre los métodos (M.) de propagación (P.) aproximada (A.) en sistemas (S.) expertos (E.).  
Envíalo a través de **"M. P. A. S. E."**.
2. En un documento en Word elabore un informe académico sobre la propagación simbólica de evidencia.  
Envíalo a través de **"Propagación Simbólica"**.

# Autoevaluación

- 1) **Un subconjunto de variables cuyos valores son conocidos, en una situación dada, se conoce como:**
  - a. Conjunto de variables.
  - b. Evidencia almacenada.
  - c. Conjunto de evidencia.
  - d. Evidencia inteligente.
  - e. Evidencias asociadas.
  
- 2) **El método de propagación en poli árboles es válido solamente para redes de estructura:**
  - a. Variable.
  - b. Compuesta.
  - c. Asociada.
  - d. Simple.
  - e. Múltiple.
  
- 3) **Dos de los métodos de propagación más importantes para este tipo de redes múltiplemente conexas son los denominados métodos de:**
  - a. Condicionamiento y agrupamiento.
  - b. Consistencia y contingencia.
  - c. Agrupamiento y consistencia.
  - d. Condicionamiento y contingencia.
  - e. Agrupamiento y consolidación.
  
- 4) **No todos los parámetros son libres, es decir, uno cualquiera de ellos puede ser escrito como:**
  - a. La unidad más la suma del resto.
  - b. La unidad por la suma del resto.
  - c. La unidad entre la suma del resto.
  - d. El total menos la suma del resto.
  - e. La unidad menos la suma del resto.
  
- 5) **El tratamiento con parámetros simbólicos es idéntico al tratamiento con valores numéricos, con la única diferencia de que las operaciones requeridas deben realizarse con un programa capaz de manipular símbolos en vez de números.**
  - a. Generación automática de código configurado.
  - b. Generación automática de código simbólico.
  - c. Generación automática de código numérico.
  - d. Generación automática de código automático.
  - e. Generación automática de código modificado.

- 6) Una medida de calidad,  $Q(B|S, \xi)$ , es un criterio mediante el cual se puede \_\_\_\_\_, donde  $B$  es una red Bayesiana,  $\xi$  la información “a priori”, y  $S$  un conjunto de datos.
- Modificar el conjunto de todas las redes Bayesianas posibles por su calidad.
  - Evaluar el conjunto de todas las redes Bayesianas posibles por su calidad.
  - Codificar el conjunto de todas las redes Bayesianas posibles por su calidad.
  - Ordenar el conjunto de todas las redes Bayesianas posibles por su calidad.
  - Transformar el conjunto de todas las redes Bayesianas posibles por su calidad.
- 7) Dado un conjunto de datos  $S$ , una medida de calidad  $Q(B|S, \xi)$  se dice que es equivalente en peso si asigna el mismo valor a todo par de redes Bayesianas equivalentes  $B_1$  y  $B_2$ , es decir, si  $Q(B_1 | S, \xi) = Q(B_2 | S, \xi)$ .
- Equivalencia en producto.
  - Equivalencia en peso.
  - Equivalencia en masa.
  - Equivalencia en códigos.
  - Equivalencia en recursos.
- 8) Una medida de calidad debe \_\_\_\_\_. Por ejemplo, debe asignarse \_\_\_\_\_ que conduzcan a la misma estructura de independenciam.
- Promover algunas propiedades deseables - la misma calidad a las redes.
  - Satisfacer algunas propiedades deseables - la misma calidad a las redes.
  - Evaluar algunas propiedades deseables - la misma calidad a las redes.
  - Reconocer algunas propiedades deseables - la misma calidad a las redes.
  - Dirigir algunas propiedades deseables - la misma calidad a las redes.
- 9) Es importante la propagación de evidencia porque:
- Permite obtener conclusiones cuando se dispone de nueva información.
  - Permite obtener direcciones cuando se dispone de nueva información.
  - Permite obtener evaluaciones cuando se dispone de nueva información.
  - Permite obtener deducciones cuando se dispone de nueva información.
  - Permite obtener comprobaciones cuando se dispone de nueva información.
- 10) Cuando no existe evidencia el proceso de propagación consiste en:
- Permitir transacciones bancarias.
  - Calcular las probabilidades marginales.
  - Continuar con problemas de probabilidad.
  - Medir las probabilidades.
  - Permitir problemas de errores.

## UNIDAD DE APRENDIZAJE IV:

### PROPAGACIÓN EXACTA EN REDES PROBABILÍSTICAS

La propagación de evidencia es una de las tareas más importantes de un sistema experto, pues permite obtener conclusiones cuando se dispone de nueva información. Cuando no se dispone de ninguna información, es decir, cuando no existe evidencia, el proceso de propagación consiste en calcular las probabilidades marginales

Estas probabilidades proporcionan información “a priori” sobre los distintos valores que pueden tomar las variables, Cuando se dispone de cierta evidencia, es decir, cuando se conoce un conjunto de variables que tienen asociadas los valores el proceso de propagación debe tener en cuenta estos valores para calcular las nuevas probabilidades de los nodos.

El poliárbol es uno de los modelos gráficos más simples para construir redes Bayesianas. En esta sección se presenta un algoritmo de propagación para este tipo de modelos probabilísticos la característica principal de este algoritmo es que su complejidad es lineal en el tamaño de la red (es decir en el número de nodos y aristas que la componen), a diferencia del método de fuerza bruta que requiere un numero exponencial de operaciones para realizar la propagación.

El método de propagación en poliárboles descrito en la sección anterior es válido solamente para redes de estructura simple (poli árboles), en las cuales existe un único camino entre cada par de nodos. Dos de los métodos de propagación más importantes para este tipo de redes son los denominados métodos de condicionamiento y método de agrupamiento.

- ❖ **ARQUITECTURA:** Consiste en la estructura organizacional de un sistema.
- ❖ **ATRIBUTO:** Es una parte específica de una clase. Una propiedad de un tipo identificada mediante un nombre.
- ❖ **ESTADO:** Es una condición o situación en la vida de un objeto, durante la cual satisface una condición, realiza una actividad o está esperando un evento.
- ❖ **ESTADO ACTIVO:** Es un estado con una acción interna y una o más transiciones asociadas a la finalización de la acción interna.
- ❖ **ESTADO COMPUESTO:** Es un estado compuesto por subestados.
- ❖ **EVENTO:** En el contexto de un diagrama de estado, un evento es un acontecimiento que puede disparar una transición de estados.
- ❖ **EVENTO TEMPORAL:** Es un evento que ocurre en un tiempo particular. Puede ser especificado por medio de una expresión temporal.
- ❖ **MÁQUINA DE ESTADOS FINITOS (MEF):** Es un modelo que describe los aspectos de control en los sistemas de información.
- ❖ **META-METAMODELO:** Es un modelo que define el lenguaje para expresar el metamodelo. La relación entre meta-metamodelo y metamodelo es análoga a la relación entre metamodelo y modelo.
- ❖ **METAMODELO:** Es un modelo que define el lenguaje para poder expresar un modelo.
- ❖ **MODELO:** Es una abstracción semánticamente consistente de un sistema.
- ❖ **MSC:** Es un lenguaje gráfico orientado a objetos que se usa para describir escenarios, es decir, ejecuciones concretas del sistema.
- ❖ **REDES DE PETRI:** Es un formalismo gráfico que permite especificar sistemas asincrónicos.
- ❖ **SUBESTADO:** Es un estado que es parte de un estado compuesto. Un subestado puede ser concurrente o disjunto.
- ❖ **SDL:** Permite expresar mediante máquinas de estados el funcionamiento de las clases del sistema.
- ❖ **SUBESTADO CONCURRENTE:** Es un subestado que puede tener cabida simultáneamente a otros subestados concurrentes en el mismo estado compuesto.
- ❖ **TIEMPO:** Es un valor que representa un momento en el tiempo, absoluto o relativo.
- ❖ **UML "UNIFIED MODELING LANGUAGE":** Es un lenguaje que permite especificar, construir, visualizar y documentar los elementos que componen un sistema de software intensivo.

# Fuentes de Información

## **BIBLIOGRÁFICAS:**

- ✚ **BENCHIMOL**, G, "Los Sistemas Expertos en la Empresa". Ed. Ra-Ma. 2011
- ✚ **CUENA**, J. "Inteligencia Artificial: Sistemas Expertos". Alianza Informática. 2010
- ✚ **GIARRATANO** H. "Sistemas Expertos. Principios y programación" 2008
- ✚ **HARMON**, P. "Sistemas Expertos". Díaz de Santos, S.A. 2007
- ✚ **LEONDES**, C. T. "Expert Systems. 6 Vo. Set." 2009.
- ✚ **NILSON**, R. "Ingeniería del conocimiento". McGraw-Hill. 2006
- ✚ **RICH**, E., Knight, K. "Inteligencia Artificial" McGraw-Hill. 2010

## **ELECTRÓNICAS:**

- ❖ **Construcción de Modelos Probabilísticos**  
[http://www.cs.us.es/cursos/iic-2010/Archivos/IIC%20-%20Teoria10\\_v03.pdf](http://www.cs.us.es/cursos/iic-2010/Archivos/IIC%20-%20Teoria10_v03.pdf)
- ❖ **Definición, motivación y origen de los sistemas expertos**  
<http://luisquillermo.com/cbse.pdf>
- ❖ **Sistemas Expertos y Modelos de Redes Probabilísticas**  
<http://fismat.umich.mx/~htejeda/gutierjm/BookCGH.pdf>
- ❖ **Sistemas expertos basados en probabilidad**  
<http://www.cs.us.es/blogs/iic2012/files/2012/02/IIC-Teoria7.pdf>



# Solucionario

## UNIDAD DE APRENDIZAJE 1

1. D
2. A
3. E
4. E
5. B
6. D
7. A
8. A
9. C
10. B

## UNIDAD DE APRENDIZAJE 2:

1. E
2. A
3. A
4. C
5. D
6. A
7. B
8. B
9. E
10. C

## UNIDAD DE APRENDIZAJE 3:

1. A
2. E
3. A
4. C
5. C
6. B
7. D
8. E
9. A
10. B

## UNIDAD DE APRENDIZAJE 4:

1. C
2. C
3. D
4. A
5. E
6. B
7. D
8. B
9. B
10. A